

Statistische Modelle in der Historischen Sozialforschung I: allgemeine Grundlagen - Deskriptivstatistik - Auswahlbibliographie

Sensch, Jürgen

Veröffentlichungsversion / Published Version

Themenheft / topical issue

Empfohlene Zitierung / Suggested Citation:

Sensch, J. (1995). Statistische Modelle in der Historischen Sozialforschung I: allgemeine Grundlagen - Deskriptivstatistik - Auswahlbibliographie. *Historical Social Research, Supplement*, 7, 1-255. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-285987>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

HSR

Supplement / Beiheft

No. 7 (1995)

Jürgen Sensch

Statistische Modelle in der
Historischen Sozialforschung I:
Allgemeine Grundlagen – Deskriptivstatistik –
Auswahlbibliographie

Köln
Zentrum für Historische Sozialforschung
1995

Inhaltsverzeichnis

EDITORIAL	8
Teil 1: Allgemeine Grundlagen	10
1. Wissenschaftstheoretische Grundlagen, Theorie und Empirie	10
2. Arten sozialwissenschaftlicher Forschungsdesigns	30
3. Der Forschungsprozeß: Ein zusammenfassender Überblick	36
Teil 2: Quantitative Methoden in der historischen Sozialforschung	48
1. Messung und Variablenbildung	48
1.1 Das »Strukturmodell« empirischer Forschung	48
1.2 Grundbegriffe der Statistik	51
2. Das Verhältnis von Grundgesamtheit und Stichprobe: Auswahlverfahren	57
2.1 Grundlagen der Stichprobenbildung	57
2.2 Zufallsgesteuerte Auswahlen	63
2.3 Geschichtete Zufallsauswahl	71
2.4 Die Klumpenstichprobe	74
2.5 Mehrstufige Auswahl	77
2.6 Willkürliche und bewußte Auswahlen	80
3. Datensammlung und -aufbereitung für statistische Analysen	86
3.1 Formaler Aufbau einer Datenmatrix	86
3.2 Tabellarische Darstellung univariater Häufigkeitsverteilungen	96
3.3 Grundlegende graphische Darstellungsformen	102
4. Zur Anwendung statistischer Modelle in der historischen Sozialforschung	117
4.1 Einleitung: Zum Stellenwert statistischer Modelle	117
4.2 Grundlegende Konzepte für das Verständnis statistischer Modelle auf deskriptiver Ebene	126
4.2.1 Lagemaße	126
4.2.2 Streuungsmaße	134
5. Bivariate Verteilungen, Assoziationsmaße und Modelle der proportionalen Reduktion des Vorhersagefehlers	143
5.1 Zweidimensionale Häufigkeitsverteilungen: Tabellarische und graphische Darstellung	144

5.2	Die Prozentsatzdifferenz	152
5.3	Maßzahlen auf der Basis des CHI-Quadrat-Modells	154
5.4	Modelle der proportionalen Fehlerreduktion	158
5.5	Die Analyse der Beziehung zwischen ordinalen Merkmalen	167
5.6	Bivariate Modelle auf der Grundlage von Streudiagrammen	177
5.6.1	Korrelationsanalyse	178
5.6.2	Lineare Einfachregression	183
Teil 3: Auswahlbibliographie		205
1.	Forschungsmethodologie, Wissenschaftstheorie und historische Methode	205
2.	Statistik	212
2.1	Auswahlverfahren	212
2.2	Deskriptiv- und Inferenzstatistik	213
2.3	Explorative Datenanalyse	217
2.4	Regressionsanalyse	218
2.5	Regressions- und Varianzanalyse auf der Basis des Allgemeinen Linearen Modells	219
2.6	Fehlende Werte in der Statistik	220
2.7	Multivariate Datenanalyse: Übersichten	220
2.8	Mehrebenenanalyse	222
2.9	Kohortenanalyse	223
2.10	Stichprobenselektivität	223
2.11	Klassifikationsverfahren/Clusteranalyse	224
2.12	Diskriminanzanalyse	224
2.13	Faktorenanalyse/Kanonische Korrelationsanalyse	225
2.14	Latent-Class-Analyse	226
2.15	Korrespondenzanalyse	226
2.16	Multivariate Kontingenztabellenanalyse	227
2.17	Logistische Regressionsanalyse/Modelle diskreter Entscheidungen	228
2.18	Verallgemeinerte lineare Modelle	229
2.19	Ereignisdatenanalyse	229
2.20	Kausalmodelle/Pfadanalyse	231
2.21	Analyse von Longitudinaldaten	233
2.22	Ökonometrie	234
2.23	Zeitreihenanalyse	235
2.24	Nichtparametrische Statistik	240
2.25	Grafische Gestaltung von Zahlen und statistische Grafik	240
2.26	Sammlungen	241
ANHANG		250

Abbildungsverzeichnis

Teil 1: Allgemeine Grundlagen

1. Wissenschaftstheoretische Grundlagen, Theorie, Empirie

- Abb. 1.1: Das Hempel Oppenheim-Schema
- Abb. 1.2: Grundlegende Wissenschaftsschemata
- Abb. 1.3: Die Struktur einer Theorie
- Abb. 1.4: Empirische Theorie und Wirklichkeit
- Abb. 1.5: Verhältnis Theorie und Empirie

3. Der Forschungsprozeß

- Abb. 3.1: Der Forschungsprozeß

Teil 2: Quantitative Methoden in der empirischen Sozialforschung

1. Messen und Variablenbildung

- Abb. 1.1: »Strukturmodell« für nicht-experimentelle Forschungsprojekte

2. Das Verhältnis von Grundgesamtheit und Stichprobe: Auswahlverfahren

- Abb. 2.1: Beziehungen zwischen den Bestimmungsgrößen der Stichprobenplanung
- Abb. 2.2: Grundgesamtheit und Stichprobe
- Abb. 2.3: Vereinfachtes Urnenschema
- Abb. 2.4: Klassifikation einfacher Techniken der Zufallsauswahl
- Abb. 2.5: Geschichtete Zufallsstichprobe als 'paralleles Urnenschema'
- Abb. 2.6: Das Auswahlmodell der Klumpenauswahl
- Abb. 2.7: Das Modell der mehrstufigen Auswahl
- Abb. 2.8: Zusammenfassung der Typen von Auswahlverfahren

3. Datensammlung und -aufbereitung für statistische Analysen

- Abb. 3.1: Allgemeines Schema für eine Tabelle
- Abb. 3.2: Grundtypen von Präsentationsgraphiken
- Abb. 3.3: Darstellung von Prozentanteilen als Flächenproportionen
- Abb. 3.4: Unterschiedliche graphische Möglichkeiten zur Veranschaulichung von Prozentanteilen
- Abb. 3.5: Säulendiagramm für den Vergleich von zeitlich geordneten Beobachtungen
- Abb. 3.6a: Kurvendiagramm: Die nationale Verschuldung in England von 1688 bis 1800
- Abb. 3.6b: Natürliche Bevölkerungsbewegung in der Bundesrepublik Deutschland und in der DDR 1946 bis 1988 (je 1000 der Bevölkerung)
- Abb. 3.7: Beispiel für einen Rangfolge-Vergleich
- Abb. 3.8: Fiktive Wertetabelle zweier metrischer Merkmale Y und X, dargestellt als zweidimensionale Punktwolke in einem Streudiagramm mit einer Geradenfunktion

- Abb. 3.9: Histogramm mit Häufigkeitspolygon
- Abb. 3.10: Exemplarische Modelltypen für Häufigkeitsverteilungen
- Abb. 3.11: Aufwärts- und abwärtskumulierte Prozentanteile

4. Zur Anwendung statistischer Modelle in der empirischen Sozialforschung

- Abb. 4.1: Die Stellung der Statistik in der historischen Sozialforschung
- Abb. 4.2: Deduktives versus induktives Vorgehen
- Abb. 4.3: Verhältnis von Daten-Theorie-Modell
- Abb. 4.4: Stellenwert statistischer Modelle
- Abb. 4.5: Prozeß der empirischen Erkenntnisbildung
- Abb. 4.6: Graphische Veranschaulichung der Lagemaße
- Abb. 4.7: Das Verhältnis der Lagemaße zueinander für verschiedene Häufigkeitsverteilungen
- Abb. 4.8: Gleicher Mittelwert zweier Verteilungen bei unterschiedlicher Streuung
- Abb. 4.9: Bestimmung des 1. und 3. Quartils aus einer Summenfunktion
- Abb. 4.10: Box-Diagramme im Vergleich
- Abb. 4.11: Box-Plots zu ausgewählten Sozialstrukturmerkmalen auf Wahlkreisebene

5. Bivariate Verteilungen, Assoziationsmaße und Modelle der proportionalen Reduktion des Vorhersagefehlers

- Abb. 5.1a: Gruppiertes Säulendiagramm der Häufigkeitsverteilung zweier Merkmale
- Abb. 5.1b: Gestapeltes Säulendiagramm zur strukturellen Darstellung einer zweidimensionalen Häufigkeitsverteilung
- Abb. 5.2: Illustration von erklärter und nichterklärter Abweichung im Rahmen der einfachen Varianzanalyse
- Abb. 5.3: Streudiagramme mit idealtypischen Punktwolken
- Abb. 5.4: Beispiel für ein Streudiagramm auf Wahlkreisebene
- Abb. 5.5: Streudiagramm mit seinem Schwerpunkt
- Abb. 5.6: Streudiagramme zur Veranschaulichung unterschiedlich hoher Korrelationskoeffizienten
- Abb. 5.7: Beispiele für lineare Funktionen
- Abb. 5.8: Deterministisches und stochastisches Regressionsmodell
- Abb. 5.9: Hypothetische Regressionsgerade und Residualgröße
- Abb. 5.10: Idealtypische Geradenverläufe
- Abb. 5.11: Katholikenanteil und NSDAP-Wahlerfolg bei der Reichstagswahl vom Juli 1932 (Freihand-Regressionsgerade)
- Abb. 5.12: Illustration unterschiedlicher Modellanpassungsgüten
- Abb. 5.13: Streuungserlegung im Regressionsmodell
- Abb. 5.14: Illustration unterschiedlich hoher Bestimmtheitsmaße
- Abb. 5.15: Unterschiedliche Modellanpassungsgüten bei gleichen Regressionskoeffizienten

Tabellenverzeichnis

Teil 2: Quantitative Methoden in der Historischen Sozialforschung

1. Messung und Variablenbildung

Tab. 1.1: Überblick über die Meßniveaus

2. Das Verhältnis von Grundgesamtheit und Stichprobe: Auswahlverfahren

Tab. 2.1: Zufallszahlen

Tab. 2.2: Zur Repräsentativität von Auswahlverfahren

3. Datensammlung und -aufbereitung für statistische Analysen

Tab. 3.1a: Formale Struktur einer Datenmatrix

Tab. 3.1b: Beispiel einer Datenmatrix: Daten zur Sozialstruktur der Wahlkreise in der Weimarer Republik aus den Volkszählungen 1925 und 1933

Tab. 3.2a: Häufigkeitsverteilung eines diskreten Merkmals

Tab. 3.2b: Die Wahl zur Nationalversammlung 1919 im Wahlkreis Bremen: Stimmanteile absolut und in Prozent der gültigen Stimmen

Tab. 3.3a: Häufigkeitsverteilung eines klassifizierten Merkmals

Tab. 3.3b: Gemeindegroßenklassen in der Weimarer Republik

Tab. 3.4: Grundtypen von Vergleichen und Graphikformen

Tab. 3.5a: Allgemeine Tabelle für eine Aufwärtskumulierung von absoluten und relativen Häufigkeiten

Tab. 3.5b: Zahlenbeispiel für aufwärts- und abwärtskumulierte Prozentanteile

4. Zur Anwendung statistischer Modelle in der historischen Sozialforschung

Tab. 4.1: Median und Mittelwert zu ausgewählten Strukturmerkmalen auf Wahlkreisebene

Tab. 4.2: Mögliche Lagemaße und ihr Anwendungsspektrum

Tab. 4.3: Streuungsmaße zu ausgewählten Sozialstrukturmerkmalen auf Wahlkreisebene

Tab. 4.4: NSDAP – Wachstum bei den Reichstagswahlen 1930 bis 1933: deskriptive Maßzahlen der univariaten Verteilungen auf Wahlkreisebene

5. Bivariate Verteilungen, Assoziationsmaße und Modelle der proportionalen Reduktion des Vorhersagefehlers

Tab. 5.1a: Notation für eine allgemeine bivariate Kontingenztafel für absolute Häufigkeiten

Tab. 5.1b: Die regionale Verteilung der Wahlkreise tabelliert gegen die klassifizierten NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (Angaben absolut)

- Tab. 5.2a: Notation für eine allgemeine bivariate Kontingenztafel mit bedingten relativen Häufigkeiten des Merkmals Y
- Tab. 5.2b: Die regionale Verteilung der Wahlkreise gegen die klassifizierten NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (relative Häufigkeiten bezüglich des NSDAP-Anteils)
- Tab. 5.2c: Die regionale Verteilung der Wahlkreise gegen die klassifizierten NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (relative Häufigkeiten bezüglich der Großregionen)
- Tab. 5.3a: Allgemeine Nomenklatur für die Zellenhäufigkeiten einer 2×2 -Tabelle
- Tab. 5.3b: Ausgewählte Regionen und dichotomisierter NSDAP-Stimmenanteil bei den Reichstagswahlen von 1933
- Tab. 5.4a: Indifferenztabelle der absoluten und erwarteten Häufigkeiten unter der Unabhängigkeitsannahme
- Tab. 5.4b: Ausgewählte Regionen und NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (Quartile)
- Tab. 5.5a: Allgemeine Mittelwerttafel als Grundlage für eine einfache Varianzanalyse
- Tab. 5.5b: Durchschnittliche NSDAP-Stimmenanteile in ausgewählten Regionen bei den Reichstagswahlen 1933
- Tab. 5.6: Variationszerlegungstabelle für die regionale Verteilung der NSDAP-Stimmenanteile bei der Reichstagswahl 1933
- Tab. 5.7a: Fiktives Zahlenbeispiel zur Illustration von konkordanten und diskordanten Paaren
- Tab. 5.7b: Ergebnisse der Paarvergleiche (fiktive Daten)
- Tab. 5.8a: Beispiel für eine Verknüpfung des X-Merkmals
- Tab. 5.8b: Mögliche Ergebnisse von Paarvergleichen in Kontingenztabellen
- Tab. 5.9a: Allgemeine 2×3 - Tabelle
- Tab. 5.9b: Ermittlung der Paare in einer allgemeinen 2×3 - Tabelle
- Tab. 5.10: Die Verteilung der Paare in einer 2×2 - Tabelle mit fiktiven Zellenbesetzungen
- Tab. 5.11: Katholikenanteil und NSDAP-Stimmenanteil bei den Reichstagswahlen von 1933
- Tab. 5.12: Übersicht zu den Assoziationsmaßen bivariater Tabellenanalysen
- Tab. 5.13: Korrelationen zwischen Katholikenanteil und NSDAP-Stimmenanteil von 1930 bis 1933
- Tab. 5.14: Additive Zerlegung der Beobachtungswerte des abhängigen Merkmals
- Tab. 5.15: NSDAP-Stimmenanteile und Katholikenanteil: Ergebnisse der einfachen Regressionsanalyse

Editorial

Dieses Supplementheft wendet sich an Leser, die grundlegende Modelle der statistischen Datenanalyse und deren Einsatz bei der Bearbeitung von Problemstellungen aus dem Bereich der Historischen Sozialforschung kennenlernen wollen. Das Heft orientiert sich an der seit 1995 praktizierten Neukonzeption der Statistikausbildung in den Herbstseminaren des Zentrums für Historische Sozialforschung. Diese Neukonzeption wurde bereits in dieser Zeitschrift (HSR, Vol. 20, 1995, No. 1, S. 110–116) vorgestellt und ist zur Information im Anhang dieses Heftes noch einmal abgedruckt.

Mit diesem Heft wird die Tradition der Veröffentlichung von Vorlesungsskripten zu den Herbstseminaren fortgesetzt. Orientiert an der früheren Konzeption der Statistikausbildung in den Herbstseminaren, lagen bislang zu den Grundkursen I und II schon zwei Supplementhefte vor:

- Thome, Helmut: Grundkurs Statistik für Historiker Teil I: Deskriptive Statistik. Historische Sozialforschung Supplement/Beiheft No. 2. Köln 1989.
- Thome, Helmut: Grundkurs Statistik für Historiker Teil II: Induktive Statistik und Regressionsanalyse. Historische Sozialforschung Supplement/Beiheft No. 3. Köln 1990.

Diese Supplementhefte haben sich in den zurückliegenden Jahren auf das Beste bewährt; allein die hohen Auflagen bzw. die umfangreichen Nachdrucke dokumentieren nachdrücklich den Erfolg beider Hefte. Auch wenn die Neukonzeption des Herbstseminars entsprechende neustrukturierte Skripten notwendig gemacht hat, werden dadurch die beiden Hefte von Helmut Thome nicht überflüssig, sondern sind als komplementär zu verstehen und finden selbstverständlich weiterhin Verwendung in den Herbstseminaren.

Mit der nun vorliegenden Einführung in die quantitative Analyse empirischer Daten sollen grundlegende Methoden der beschreibenden Statistik und deren Anwendungsvoraussetzungen erläutert werden. Ferner soll das Skript in die Probleme des Verhältnisses von formalen Modellen und statistischen Konzepten einführen. Die Anwendung der statistischen Modelle wird am Beispiel eines Forschungsprojektes aus der Historischen Wahlforschung erläutert. Die Beispiele sollen dem Leser den Aussagegehalt statistischer Modelle und deren Anwendungsmöglichkeiten verdeutlichen.

In Teil I wird die wissenschaftstheoretische Position empirischer Forschung behandelt. Es werden grundlegende sozialwissenschaftliche Forschungsdesigns vorgestellt und ein Überblick über die einzelnen Phasen des empirischen Forschungsprozesses gegeben.

In Teil II werden zunächst Grundbegriffe der Statistik erläutert. Da die Datengewinnung für die statistischen Analysen untrennbar mit Fragen der Stichprobenbildung verbunden ist, werden auch die wichtigsten Auswahlverfahren behandelt. Nach der Datenaufbereitung können statistische Daten durch ge-

eignete graphische Darstellungen oder in Form von Tabellen veranschaulicht und durch statistische Maßzahlen verdichtet werden. An die Datenbeschreibung schließt die bivariate Datenanalyse an. Neben klassischen Methoden der Analyse von zweidimensionalen Kontingenztabellen mit Hilfe von Assoziationsmaßzahlen wird die Varianz-, Korrelations- und Regressionsanalyse behandelt.

Die Auswahlbibliographie in Teil III schließt thematisch auch schon die Literatur zum Aufbaukurs des Herbstseminars ein. Die Bibliographie enthält die grundlegenden Monographien für die in der Historischen Sozialforschung wichtigsten Gebiete der statistischen Datenanalyse. Die Literatur soll dem Leser einen einfachen Zugang zu den verschiedensten Gebieten ermöglichen.

Statistische Analyseverfahren sind wegen ihrer umfangreichen und aufwendigen Rechenabläufe erst mit der Verbreitung der EDV ein umfassend angewandtes Instrument der empirischen Forschung geworden. Aufgrund einer in den letzten Jahren sprunghaft gewachsenen Verfügbarkeit unterschiedlichster Programmpakete und Versionen – insbesondere im Bereich der Personal Computer – lassen sich heute auch umfangreiche und aufwendige statistische Datenanalysen innerhalb kürzester Zeit ausführen. Da die Statistik-Softwarepakete einem raschen Wandel unterliegen und daher ihre Beschreibung einer ständigen Aktualisierung bedarf, erschien es sinnvoll, einen vergleichenden Überblick über die unterschiedlichen statistischen Auswertungssysteme vorab zu publizieren (Sensch, Jürgen: Der Einsatz von Statistik-Programmpaketen: Klassifikation und vergleichender Überblick, in: HSR, Vol. 20, Heft 3, Special Issue: PC-Software für die statistische Analyse in der Historischen Sozialforschung, S. 3 – 62).

In einem zweiten Heft »Statistische Modelle in der Historischen Sozialforschung II«, das für 1997 geplant ist, werden die Lerninhalte des Aufbaukurses des Herbstseminars, insbesondere multivariate Analyseverfahren dargestellt. Schwerpunktmäßig sollen folgende Themen behandelt werden: Wahrscheinlichkeitstheorie und Inferenzstatistik, Drittvariablenkontrolle und Partialität, Überblick über die Anwendung von multivariaten Analyseverfahren, das Allgemeine lineare Modell (multiple Regressions- und Varianzanalyse), Diskriminanzanalyse, Kanonische Korrelationsanalyse, Faktorenanalyse, Analyse linearer Kausalmodelle, der LISREL-Ansatz der Kausalanalyse, mehrdimensionale Kontingenztabellenanalyse, Korrespondenzanalyse, Clusteranalyse, Ereignisdatenanalyse.

Es ist mir eine angenehme Pflicht meinen KollegInnen Evelyn Brislinger, Heiko Fauck, Dr. Rainer Metz, Ralph Ponemereo, PD Dr. Wilhelm H. Schröder und Dr. Jürgen Wilke für viele kritische Kommentare und Verbesserungsvorschläge zu danken. Sämtliche verbliebenen Fehler gehen zu Lasten des Autors, der für kritische Hinweise jederzeit dankbar ist.

Köln, im Dezember 1995

Jürgen Sensch

Teil 1: Allgemeine Grundlagen

1. Wissenschaftstheoretische Grundlagen, Theorie und Empirie

Im allgemeinsten Sinne kann *Wissenschaft* als eine erkenntnisbezogene, geistige Tätigkeit verstanden werden. Als erkenntnisbezogene Tätigkeit verfolgt die Wissenschaft das Ziel, die Welt systematisch und analytisch zu erfassen und das uns zur Verfügung stehende Informationsspektrum mittelbar bzw. unmittelbar zu erhöhen. Dieses Ziel wird durch die Konstruktion und die Überprüfung von Theorien und Hypothesen einerseits und deren Anwendung zur Beschreibung, Erklärung und zur Voraussage von Phänomenen innerhalb der realen Welt andererseits angestrebt. 'Erklärung' und 'Voraussage' können als die Grundfunktionen des wissenschaftlichen Denkens bezeichnet werden. Neben dieser Konzeptualisierung von Wissenschaft sind selbstverständlich auch andere Formen von 'Wissenschaften' konstruierbar, der hier zugrundegelegte Terminus schränkt die Konzeptualisierung jedoch auf 'empirische Wissenschaft' ein: Es sollte von keinen Postulaten (wissenschaftliche Sätze) ausgegangen werden, die aus logischen Gründen nicht von jedem Wissenschaftler empirisch (d.h. durch Erfahrung) überprüfbar sind.

Der Begriff *Empirie* leitet sich aus dem Griechischen her und bedeutet 'Sinnerfahrung'. *Empirische Wissenschaft* ist demnach der Teil der Wissenschaften, der auf der Erfahrung durch die menschlichen Sinne beruht. Empirisches Vorgehen ist »Ausgehen von Erfahrungstatsachen«. Empirische Wissenschaft wird daher üblicherweise Erfahrungswissenschaft genannt. Bei der Erfassung von »Erfahrungstatsachen« bedient man sich bestimmter Techniken wie Beobachtung, Experiment, Befragung, Akten-/Dokumentenanalyse usw. Dies ist nicht gleichbedeutend mit theorielosem Vorgehen, vielmehr ist empirische Forschung ohne Rückgriff auf Theorien (explizite oder zumindest 'Alltagstheorien') gar nicht möglich. Wir werden diese Sichtweise weiter unten noch ausführlicher verdeutlichen.

Damit sind bereits zwei zentrale Kennzeichen von empirischer Wissenschaft genannt: Überprüfbarkeit und Intersubjektivität. *Intersubjektive Überprüfbarkeit* bedeutet, daß innerhalb einer Wissenschaftlergemeinschaft zwischen einzelnen Forschern Einigkeit über den Sinn wissenschaftlicher Sätze sowie über die Methoden des Beweises und der Überprüfbarkeit der Aussagen hergestellt werden kann. Die theoretischen Vorstellungen müssen gemäß den geltenden wissenschaftlichen Maßstäben und Regeln formuliert und überprüft werden, um den qualitativen Sprung von einer alltagsweltlichen Theorie zu einer wissenschaftlichen Theorie zu erreichen. Unser alltägliches Common-Sense-Denken ist durch ein ganz pragmatisches, auf die Bewältigung immer wieder auf-

tretender Situationen gerichtetes Interesse gekennzeichnet; in ihm mischen sich persönliche Erfahrungen, Vorurteile, wissenschaftliche Einzelergebnisse, Hypothesen und Wertungen in großer Vielfalt. Wissenschaftliches Denken aber sollte einem erkenntnistheoretischen System folgen: nach Regeln vorgehen, Vermutungen empirisch überprüfen und für andere nachvollziehbar (und nachprüfbar) machen. Die Inhalte wissenschaftlicher Sätze geben nicht mehr widerspruchsfrei die subjektiven Einschätzungen bzw. 'Meinungen' des jeweiligen Forschers wieder. Denn nach unserer Definition von 'Wissenschaft' müssen die wissenschaftlichen Sätze objektive, d.h. empirisch überprüfbare Aussagen beinhalten, die stets den Charakter von überprüfbaren Hypothesen haben.

Als *Methode* wird allgemein ein System von Regeln zur Erreichung eines bestimmten Ziels bezeichnet, das ein planmäßiges und systematisches Vorgehen ermöglicht. Nach diesen Regeln werden Instrumente verwendet. Die Aufstellung von *Verfahrensregeln* heißt die Spezifizierung allgemeiner Methodenregeln auf den jeweils speziellen Anwendungsfall. Als erprobte Problemlösungsmittel entlasten sie den Forscher bei empirischen Untersuchungen. Sie sind allerdings nicht einfach routinemäßig zu verwenden, sondern müssen dem jeweiligen Gegenstandsbereich flexibel angeglichen werden. Wissenschaft als Handeln zu begreifen, das auf die Herstellung von wissenschaftlichen Aussagen mit den oben festgelegten Kriterien zielt, lenkt den Blick auf die Regeln, die diesem Handeln zugrundegelegt werden müssen. Die Regeln beschreiben im allgemeinen die Wissenschaftstheorie und im besonderen die jeweilige fachwissenschaftliche Methodologie (Methodenlehre).

Die *empirische Methodenlehre* beinhaltet demnach Methoden und Verfahrensweisen, die der Gewinnung von Erfahrungen auf der Grundlage planmäßigen und systematischen Vorgehens dienen. Beschränkungen unserer Erfahrungsmöglichkeiten, die sich in der Alltagserfahrung aus der fehlenden Systematik, der Begrenztheit des menschlichen Sensoriums u.ä. ergeben, sollen hiermit bis zu einem gewissen Grade überwunden werden. Wissenschaftliches Vorgehen besteht darin, den Erkenntnisprozeß bewußt, systematisch, methodisch reflektiert und kontrolliert voranzutreiben. Empirische Forschung im engeren Sinn sammelt dazu notwendige Erfahrungen mit Hilfe der Regeln, Verfahrensweisen und Techniken, wie sie die empirische Methodenlehre bereitstellt.

Mit dem Terminus *Wissenschaftstheorie* (Theorie über Wissenschaft) bzw. mit den Begriffen *Logik der Forschung* ('Philosophy of Science') wird jene kritische Untersuchung des wissenschaftlichen Denkprozesses sowie jene Analyse wissenschaftlicher Ergebnisse bezeichnet, die eine rationale Ausformung der wissenschaftlichen Forschungsergebnisse fördern, d.h. das Nachdenken über die Entstehung, die Überprüfung und die Strukturierung von wissenschaftlichen Theorien. *Wissenschaftstheoretische Leitvorstellungen* (d.h. Vorstellungen über die Weise, wie Wissenschaft betrieben werden sollte) bestimmen das wissenschaftliche Handeln dadurch, daß sie einen Rahmen abstecken,

innerhalb dessen Forschung (und Lehre) stattfinden. Die obengenannten Kriterien 'Intersubjektivität' und 'empirische Überprüfbarkeit' sind Bestandteile eines bestimmten Rahmens.

Es gibt unterschiedliche wissenschaftstheoretische Regelwerke, welche um die Anerkennung in der Wissenschaftlergemeinschaft ringen. Die Wissenschaftstheorie gibt es nicht, eher findet man mehrere unterschiedliche Positionen (oder 'Schulen') mit unterschiedlichen Auffassungen über wissenschaftliches Vorgehen. Entsprechend dieser Vieldimensionalität in der Sicht auf Wissenschaft, fallen auch die Aussagen über die Aufgaben und Ziele von Wissenschaften sehr vielfältig aus. Es geht – sehr vereinfacht ausgedrückt – um die Frage, ob Wissenschaften *beschreiben* (Positivismus), *erklären* (Kritischer Rationalismus), *verstehen* (Hermeneutik) oder *verändern* (Kritische Theorie) sollen. Je nachdem wie die Antworten ausfallen, ergeben sich weitreichende Konsequenzen für den wissenschaftlichen Arbeitsbetrieb.

Eine bedeutsame forschungslogisch-methodologische Version erfahrungswissenschaftlicher Vorgehensweise ist unter dem Namen *Kritischer Rationalismus* ausformuliert, als deren Hauptvertreter K.R. Popper (1902–1994) und in Deutschland H. Albert (geb. 1921) zu gelten haben. Der Kritische Rationalismus ist weitgehend die wissenschaftstheoretische Grundlage empirischer Forschung. Das Hauptprinzip empirischer Forschungsmethodologie – wie sie vom Kritischen Rationalismus vertreten wird – lautet: Aussagen der Erfahrungswissenschaften sollen über die Realität eines Gegenstandsbereiches, für den sie aufgestellt wurden, informieren (d.h. die verwendeten Begriffe müssen sich auf die erfahrbare Realität beziehen). Zentrale Aufgabe eines Wissenschaftlers (mit erfahrungswissenschaftlicher Orientierung) ist es, Theorien über die Realität eines Gegenstandsbereiches aufzustellen und diese zu überprüfen. Sämtliche Aussagen müssen an der Erfahrung überprüfbar sein, d.h. müssen sich in der Konfrontation mit der Realität »intersubjektiv«, d.h. unter gegenseitiger Kontrolle innerhalb der 'scientific community' durch den kritischen Diskurs, überprüfen lassen. In den Wissenschaften soll die »Methode der rationalen und kritischen Diskussion« vorherrschend sein. Nach Popper: »Sämtliche Aussagen einer empirischen Wissenschaft müssen prinzipiell an der Erfahrung scheitern können« (vgl. Popper, K.R., 1971: Logik der Forschung, Tübingen, S. 15). Das Ziel der Wissenschaften soll nicht länger darin bestehen, Theorien (d.h. Sätze) als wahr (verifizierbar) zu erweisen; vielmehr sollen Theorien ständigen Widerlegungsversuchen (Falsifikationen) ausgesetzt werden. Das hat drei Konsequenzen für den Geltungsbereich so abgegrenzter erfahrungswissenschaftlicher Aussagen:

- Nur solche Begriffe können in erfahrungswissenschaftlichen Aussagen verwendet werden, die sich auf die erfahrbare Realität beziehen (empirischer Bezug der verwendeten Begriffe);
- die formulierten Sätze oder Aussagen empirischer Wissenschaft müssen eine Beschreibung von Sachverhalten bieten, die ebenfalls prinzipiell erfahrbar sind (empirischer Bezug der Gesamtaussage);

- die Sätze müssen so formuliert sein, daß sie prinzipiell widerlegbar sind. Als empirische Aussagen nicht zugelassen sind daher analytisch wahre, d.h. aus logischen Gründen wahre Aussagen (z.B. Tautologien; Beispiel: »Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter, oder es bleibt wie es ist«).

Die Verifikation von Theorien ist nach Popper nur »wahrheitskonservierend«, sie treibt die Wissenschaft nicht zu immer neuen Lösungsversuchen von Problemen. Grundlegend für Poppers Wissenschaftslogik ist die Überlegung, daß der Schluß von einzelnen Erfahrungen auf allgemeine Theorien logisch unzulässig ist. Auch wenn man noch so viele empirische Untersuchungen durchführt, so lassen sich aus diesen Einzelbeobachtungen nie logisch zwingend allgemeine Sätze (Theorien) schließen. Auf dieses sog. *Induktionsproblem* (Induktion: Schließen vom Einzelnen auf das Allgemeine) hat schon der Philosoph David Hume hingewiesen. In Theorien werden generelle Aussagen angestrebt. Die zentrale Frage dazu lautet: Wie können die sehr allgemeinen, uneingeschränkten Behauptungen, aus denen sich unsere Theorien zusammensetzen, auf der Grundlage einer nur begrenzten Anzahl von Beobachtungen gerechtfertigt werden? Hume bestreitet das Ableitungsverfahren der Induktion mit dem Argument, daß keine noch so große Zahl von Beobachtungsaussagen die Wahrheit einer echten Allaussage rein logisch begründen kann. Denn Induktion ist immer ein Erweiterungsschluß, man schließt auf etwas, was in den Annahmen nicht enthalten ist. Wenn man die Beobachtung macht, daß bei einer Gelegenheit auf das Ereignis A das Ereignis B folgt, dann ergibt sich daraus nicht die logische Folgerung, daß B auch bei einer anderen Gelegenheit auf A folgt. Ein solcher Schluß folgt weder aus hundert noch aus tausend Beobachtungen dieser Art.

Popper glaubt, dieses Induktionsproblem gelöst zu haben (Popper, K.R., 1973: Objektive Erkenntnis. Ein evolutionärer Entwurf. Hamburg, 1973, S. 13), indem er zwar der Erkenntnis zustimmt, daß sich aus noch so vielen Beobachtungen nie die Wahrheit einer allgemeinen Theorie rechtfertigen läßt, er aber hinzufügt, daß sich auf Grund der logischen Asymmetrie zwischen Verifikation und Falsifikation eine Theorie durch einzelne Beobachtungen zwar nicht als wahr, wohl aber als falsch erweisen kann. Diese logische Asymmetrie wird an dem so bekannt gewordenen Beispiel Poppers von den weißen und schwarzen Schwänen deutlich: angenommen man hat bisher in seinem Leben nur weiße Schwäne gesehen und bildet aufgrund dieser Erfahrungsbasis mittels eines Induktionsschlusses den Allsatz: »Alle Schwäne sind weiß«. Dieser Allsatz bezieht sich nun nicht nur auf die beobachteten, sondern auch auf alle gegenwärtigen, in der Vergangenheit und in der Zukunft existierenden Schwäne. Daraus folgt, daß wir uns dieser Aussage: »Alle Schwäne sind weiß« eigentlich nie sicher sein können (bzw. sie nie endgültig verifiziert werden kann), denn es könnte immer ein Zeitpunkt kommen oder ein Fleck auf dieser Erde entdeckt werden, an dem ein Schwan beobachtet wird, der nicht weiß ist. Dagegen reicht

die Existenz eines einzigen schwarzen Schwans aus, um den Satz »Alle Schwäne sind weiß« zu widerlegen (zu falsifizieren). Nach kritisch-rationaler Auffassung können universale, d.h. räumlich und zeitlich unbeschränkt geltende Gesetzesaussagen niemals aufgrund vergangener und gegenwärtiger Erfahrungen verifiziert werden, weil sie sich auch auf künftige Ereignisse beziehen, die sich hier und jetzt empirisch nicht erfassen lassen.

Theorien bleiben für Popper auch nach strengen Prüfungen immer fehlbar und vorläufiges Wissen. Theorien können sich zwar bewähren, aber niemals den Status eines endgültig gesicherten Wissens erlangen. Der Fortschritt der Wissenschaft läßt sich charakterisieren als ein Fortschritt zu Theorien von immer größerer *Wahrheitsähnlichkeit*. In diesem Zusammenhang plädiert Popper dafür, in der Wissenschaft die Idee der Wahrheit im Sinne einer regulativen Idee aufzufassen (*Approximationstheorie der Wahrheit*): Theorien bleiben grundsätzlich immer hypothetisch, aber das Streben nach Wahrheit soll den Wissenschaftler immer weiter in seinen kritischen Bemühungen vorantreiben.

Diese Aufforderung an den Wissenschaftler, seine eigenen Erwartungen ständig zu korrigieren und mit Widerlegungsversuchen zu konfrontieren, ist eine große Anforderung; denn für jeden Menschen ist es aus seinem alltäglichen Leben heraus naheliegender und auch selbstwertschützender, nach Bestätigungen seiner Theorien zu suchen, als seine Forschungen ständig in Frage zu stellen. Eine besondere Funktion erhält an dieser Stelle der Gedanke der intersubjektiven Kritik. Wenn der Forscher selbst es vielleicht nicht schafft, seine Theorie wirklich kritisch zu prüfen, dann können immer noch die anderen Forschungskollegen (vor allem die mit einer konträren theoretischen Auffassung) versuchen, die Theorie zu widerlegen. Es gilt als wesentliches Moment der Popperschen Forschungslogik, an der Forderung nach einer Qualifizierung von Wissenschaftlern zu rationaler und kritischer Diskussionsfähigkeit festzuhalten. Wissenschaft ist keine individuelle Angelegenheit der einzelnen Wissenschaftler, »sondern eine soziale Angelegenheit ihrer gegenseitigen Kritik« (Popper, a.a.O., S. 112).

Die Verwirklichung der intersubjektiven Nachprüfbarkeit setzt vor allem einen Konsens der Wissenschaftler über bestimmte allgemeine *Spielregeln* voraus. Die »Idee der gegenseitigen rationalen Kontrolle durch kritische Diskussion« (Popper, a.a.O., S. 18) kommt hauptsächlich dann zum Tragen, wenn Regeln, die eine präzise Inhaltlichkeit und Fruchtbarkeit der Sprache gewährleisten, entwickelt und befolgt werden. Nur dann ist garantiert, daß Aussagen von anderen Wissenschaftlern tatsächlich verstanden und deshalb auch nachgeprüft werden können. Das Ziel der gegenseitigen und mithin objektiven Kritik ist die Eliminierung oder Modifikation von Aussagen, die sich als falsch erwiesen haben. Eine Wissenschaft, die einer ständigen Wahrheitssuche verpflichtet ist und hierbei gemäß dem Trial-and-error-Prinzip, der Methode von Versuch und Irrtum, verfährt, läßt als Beweismaterial nicht irgendwelche bestätigenden empirischen Daten gelten, sondern nur jene empirischen Belege

und Argumente, die zu den jeweiligen Hypothesen im Widerspruch stehen. Die Hypothese, daß alle Schwäne weiß sind, kann niemals endgültig bestätigt werden, da nicht alle Zeit-Raum-Gebiete abgesucht werden können, um die Behauptung zu prüfen; aber sie kann bereits durch einen schwarzen Schwan, der gefunden wird, widerlegt werden. Nur durch die Widerlegung unserer Irrtümer bekommen wir somit Kontakt mit der Realität. Folgendes Vorgehen wird daher von den Kritischen Rationalisten bei der Überprüfung empirischer Aussagen bei dem Herantasten an die (nicht endgültig beweisbare) Wahrheit empfohlen: Hat sich eine Hypothese oder eine Theorie als falsch erwiesen, dann wird diese Hypothese/Theorie verworfen. Sie ist damit nicht hinfällig, aber sie kann in der gegenwärtigen Formulierung keine Geltung mehr beanspruchen. Sie muß unter Berücksichtigung der neugewonnenen Erkenntnisse umformuliert werden, so daß ihr »Falschheitsgehalt« eliminiert wird. Diese neue Theorie oder Hypothese ist wiederum einer empirischen Überprüfung zu unterziehen. Wird sie wieder falsifiziert, ist sie erneut zu modifizieren und empirisch zu untersuchen usw.

Die Falsifikation ist ebenso wie die Verifikation abhängig von der Existenz bestimmter »Beobachtungsaussagen«. Popper hält trotz aller Modifikationen an einem empiristischen Grundprinzip fest, wenn er behauptet: »Ein empirisch-wissenschaftliches System muß an der Erfahrung scheitern können« (Popper, a.a.O., S. 15). Die Sätze, die eine Theorie scheitern lassen können, also die Vermittlungssätze zwischen den Hypothesen und der »Wirklichkeit« sind im Kritischen Rationalismus singuläre »Es gibt«-Sätze und heißen *Basissätze*. Sie machen Aussagen über Ereignisse in Raum und Zeit, die beobachtbar sind und auch durch Experimente gewonnen werden können. Allerdings geht Popper nicht mehr wie ein »naiver Empirist« davon aus, daß die Wissenschaft mit Basissätzen beginnt. Vielmehr weist Popper darauf hin, daß am Anfang jeglicher Wissenschaft Theorien stehen. Beobachtungen und damit Basissätze ergeben sich nur auf Grund von bestimmten theoretischen Interessen: »In der Beobachtung haben wir es mit einer Wahrnehmung zu tun, die planmäßig vorbereitet ist, die wir nicht 'haben' (wie eine Sinneswahrnehmung), sondern 'machen', wie die deutsche Sprache ganz richtig sagt. Der Beobachtung geht ein Interesse voraus, eine Frage, ein Problem – kurz, etwas Theoretisches« (Popper, a.a.O., S. 371). Theorien leiten unsere Informationsaufnahme, und es gibt keine Beobachtungen ohne Theorien¹.

Diese wichtige wissenschaftstheoretische Erkenntnis, daß es keine »reinen«, von Theorien, von Interessen, also vom erkennenden Subjekt unabhängigen Beobachtungen als Basis für Wissenschaft gibt (wie sehr wir uns dieses unbezweifelbare Fundament auch wünschen mögen), führte Popper konse-

¹ Popper (1973) vergleicht diese Bedeutung der Theorie für die Beobachtungen mit Scheinwerfern: Theorien lassen unsere Beobachtungen nicht nur in einem bestimmten Licht erscheinen, sondern sie leiten uns auch in unserem Voranschreiten zu immer neuen Beobachtungen.

quenterweise zu der Folgerung, daß auch Beobachtungsaussagen prinzipiell immer fehlbar sind. Wenn auch die Beobachtungsaussagen keine sichere Basis für die Überprüfung von Theorien sind, wie können dann Theorien überhaupt widerlegt (falsifiziert) werden?

Aus diesem schwierigen Problem innerhalb der Konzeption des Kritischen Rationalismus führt uns Popper mit dem sog. *Koventionalismus*: Die Basissätze werden durch Beschluß, durch Konvention anerkannt, sie sind Festsetzungen. Ein Basissatz gilt dann als (vorläufig) akzeptiert, wenn bei Einhaltung der geltenden methodologischen Regeln einer Wissenschaft innerhalb der Forschergemeinschaft Einigkeit über die Gültigkeit hergestellt werden kann (intersubjektive Einigkeit über die Erfahrbarkeit des Beobachteten).

Wann werden nach Popper Theorien als wissenschaftlich anerkannt? Theorien sind nach seinem Modell wissenschaftlich, wenn sie empirisch, intersubjektiv nachprüfbar, widerspruchsfrei, falsifizierbar und wertfrei sind. Weiterhin kennzeichnet er wissenschaftliche Theorien als allgemeine Sätze, welche Zusammenhänge zwischen verschiedenen in der Realität beobachtbaren Ereignissen behaupten. Aufgestellt werden »Wenn . . . , dann . . .«-Beziehungen zwischen Ereignissen, d.h. es werden Ursache-Wirkungs-Zusammenhänge analysiert.

Als ein Beurteilungskriterium für die Güte von Theorien und aus ihnen abgeleitete Hypothesen führte Popper den Begriff des *Informationsgehalts* ein. Informationshaltige, d.h. empirisch gehaltvolle Hypothesen sollten möglichst allgemein sein (d.h. sie sollten in möglichst vielen Realitätsbereichen Gültigkeit beanspruchen) und sie sollten möglichst präzise sein (d.h. sie sollten in vielen unterschiedlichen Situationen hinsichtlich ihrer Geltung untersucht werden). Der Allgemeinheitsgrad bezieht sich auf die *Wenn*-Komponenten; der Präzisionsgehalt auf die *Dann*-Komponenten. Mit dem Konzept des Informationsgehalts wurde von Popper ein Kriterium für wissenschaftlichen Fortschritt und damit auch ein Kriterium für den Vergleich von wissenschaftlichen Theorien aufgestellt. Je präziser und je umfassender eine Theorie ist, desto anfälliger ist sie für Falsifikationen und desto größer ist ihr empirischer Informationsgewinn und damit auch der wissenschaftliche Erkenntniszugewinn.

Bisher wurden vor allem Probleme der Bildung von wissenschaftlichen Theorien und deren Überprüfungsmöglichkeiten im Licht des Kritischen Rationalismus erörtert. Im folgenden soll der Frage nachgegangen werden, wozu wir Theorien in der Wissenschaft benötigen, denn die Fragen nach dem Ziel und Zweck von wissenschaftlicher Arbeit gehören ebenso in den Bereich der Wissenschaftstheorie wie die Fragen nach den Grundlagen wissenschaftlicher Theorien. Gefordert werden im Rahmen des Kritischen Rationalismus vor allem drei Funktionen von wissenschaftlichen Theorien:

- Theorien sollen bestimmte Ereignisse erklären (warum liegt ein bestimmter Sachverhalt vor?),

- die Folgen bestimmter Gegebenheiten vorhersagen (prognostizieren) und
- die Wege angeben, wie man bestimmte Ziele praktisch erreichen kann (Ableitung von Technologien).

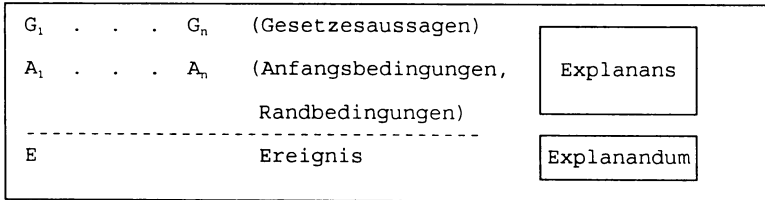


Abb. 1.1: Das Hempel Oppenheim-Schema

Quelle: Hempel, C.G./Oppenheim, P., 1948: Studies in the logic of explanation, in: Philosophy of Science, 15, S. 135-175.

Den Vorgang der Erklärung beschreibt Popper als logische Ableitung. Einen Vorgang 'kausal erklären' heißt, einen Satz, der ihn beschreibt, aus Gesetzen und Randbedingungen deduktiv ableiten. Hempel & Oppenheim (1948) haben diesen Erklärungsbegriff in ein wissenschaftstheoretisches, *deduktiv-nomologisches Beschreibungsmodell* (das sog. Hempel-Oppenheim-Schema: H-O-Schema) überführt, das in Abbildung 1.1 dargestellt ist (Hempel, C.G./Oppenheim, P., 1948: Studies in the logic of explanation, in: Philosophy of Science, 15, S. 135–175). Wichtig in diesem H-O-Schema sind vor allem zwei Begriffe: das Explanans (d.h. das, was erklärt) und das Explanandum (d.h. das, was zu erklären ist). Das Explanans setzt sich zusammen aus zwei Klassen von Sätzen: Erstens den Aussagen über allgemeine Gesetzmäßigkeiten (Hypothesen, Theorien, Naturgesetze) und zweitens den Beschreibungen der für einen besonderen Fall gegebenen Anfangsbedingungen und der Randbedingungen.

- Das *Explanandum*: Die Erklärung eines Phänomens bedeutet im Prinzip, das zu erklärende Phänomen als die *Folge* bestimmter (kausaler) Ursachen zu erkennen. Ausgangspunkt einer jeden Erklärung ist eine Aussage, die das zu erklärende Phänomen korrekt beschreibt. Dies sind meist spezielle Sachverhalte mit Angaben über Raum und Zeit des Auftretens, oft sog. singuläre Ereignisse, d.h. solche, die mit festen Angaben über Ort und Zeit des Auftretens versehen sind.
- Das *Explanans*: Die Erklärung des Phänomens besteht in dem Nachweis, daß die Aussage über das Explanandum in bestimmter Weise in einer Klasse von anderen Aussagen *logisch* enthalten ist. Diese Klasse von erklärenden Aussagen wird als Explanans bezeichnet; es hat selbst wiederum zwei Bestandteile: Allgemeine Gesetze und Randbedingungen.
- *Allgemeine Gesetze*: Das Explanans enthält Aussagen über (mindestens) ein allgemeines Gesetz. Dieses Gesetz besteht wiederum aus zwei Grundele-

menten: Aus einem Teil, in dem bestimmte Ursachen aufgeführt sind; und aus einem Teil, in dem gewisse *Folgen* benannt werden, die – laut Gesetz – immer dann auftreten, wenn die Ursachen gegeben sind. Gesetze verbinden die Ursachen mit Folgen, indem sie die Folgen als *Funktion* der Ursachen benennen. Der Ursachenteil von Gesetzen ist die *Wenn*-Komponente, der Folgeteil ist die *Dann*-Komponente des Gesetzes.

Das Gesetz benennt eine (möglichst) allgemein geltende funktionale Beziehung zwischen Ursachen und Folgen. Bedeutsam ist, daß es sich um die explizite und möglichst präzise Angabe eines funktionalen *Zusammenhangs* handelt. Oft wird diese Funktion als *Kausalbeziehung* interpretiert.

- *Randbedingungen*: Zweitens sind in dem Explanans Aussagen über die sog. Randbedingungen aufgeführt. Dabei handelt es sich – wie bei dem Explanandum – um Beschreibungen, diesmal aber darüber, daß die in der *Wenn*-Komponente des Gesetzes genannten Bedingungen auch tatsächlich vorliegen. Die Randbedingungen beziehen sich daher immer auf im Vergleich zur Geltung des Gesetzes spezielle Sachverhalte (oft ebenfalls auf sogenannte singuläre Ereignisse).

Die Erklärung eines Explanandums ist dann erfolgt, wenn es ein Gesetz gibt, das das Explanandum allgemein als Folge der Randbedingungen aufführt, und wenn gezeigt werden kann, daß die im Gesetz für diese Folgen geforderten Randbedingungen im vorliegenden speziellen Fall auch wirklich erfüllt waren.

Jede Erklärung ist ein deduktiver Schluß: Aus dem Gesetz und aus den Randbedingungen folgt das Explanandum logisch. Erklären als wichtigstes Ziel der deduktiv-nomologischen Wissenschaft heißt, daß man den zu erklärenden Sachverhalt aus Gesetzesaussagen und Rand- bzw. Anfangsbedingungen, die eine Exemplifizierung der *Wenn*- oder *Je*-Komponente des Gesetzes darstellen, deduziert. Die *Wenn*- beziehungsweise *Je*-Komponente einer Hypothese wird, weil sie bei bestimmter Betrachtung von nichts Weiterem mehr abhängt, auch als unabhängige Variable (oder Wirkfaktor bzw. Determinante) und die *Dann*- beziehungsweise *Desto*-Komponente als abhängige Variable (Resultante) bezeichnet. Definitionsgemäß besteht zwischen diesen beiden Variablen eine asymmetrische Beziehung. Hempel und Oppenheim haben insgesamt vier *Adäquatheitsbedingungen* angegeben, die bei einer angemessenen Erklärung erfüllt sein müssen:

- Das Explanandum muß im Explanans tatsächlich logisch enthalten sein.
- Das Explanans muß (mindestens) ein allgemeines Gesetz enthalten.
- Das Explanans muß empirischen Gehalt besitzen, d.h. Gesetz und Randbedingungen müssen empirisch prüfbar sein.
- Die Aussagen im Explanans müssen wahr sein.

Diese Bedingungen verweisen auf die nicht zu unterschätzende Bedeutung valider Beschreibungen schon für den Ausgangspunkt von Erklärungen. Und damit: Auf die Wichtigkeit aller dafür entwickelten Methoden der Datenerhebung und statistischen Analyse.

So übersichtlich dieses Schema auch ist, so muß doch ausdrücklich betont werden, daß die Übertragung auf die sozialwissenschaftliche Forschungswirklichkeit nicht ohne entscheidende Veränderungen und Einschränkungen möglich ist. Das H-O-Schema geht von sogenannten *deterministischen Aussagen* mit einem universellen und unbeschränkten Geltungsanspruch in den Wissenschaften aus. Deterministische Aussagen behaupten: »Wenn X, dann immer Y«. Solche streng nomologischen Aussagen lassen sich allenfalls in den Naturwissenschaften empirisch nachweisen. Problematisch (und deshalb weit schwieriger als bisher dargestellt) ist es jedoch für die Sozialwissenschaften, da diese kaum über deterministische Gesetze verfügen (wie z.B. »Alle Menschen zeigen ein Verhalten zur Wiederherstellung von Freiheit, wenn man ihre Freiheit einengt«). Verfügbar sind meist nur sog. probabilistische Gesetze mit einem nicht-deterministischen Anspruch (z.B. »75 Prozent aller Menschen, welche sich in ihrer Freiheit eingeengt fühlen, zeigen in bestimmten Situationen ein Verhalten, ihre Freiheit wieder herzustellen«). Sozialwissenschaftliche Gesetze zeichnen sich wesentlich durch *Wahrscheinlichkeitsgrade* aus. Für den sozialen Bereich mit seinen besonders »komplexen Phänomenen« (vgl. von Hayek, F.A., 1972: Die Theorie komplexer Phänomene, Tübingen) stehen in der Regel nur statistische (d.h. probabilistische) Gesetzesaussagen zur Verfügung. Probabilistische Aussagen behaupten: »Wenn X, dann mit a%-Wahrscheinlichkeit Y«. Das Auftreten eines Ereignisses ist nur mit einer bestimmten (empirisch zu ermittelnden, d.h. induktiven) Wahrscheinlichkeit zu erwarten.

Daß wir in den Sozialwissenschaften nicht über deterministische Gesetze verfügen, könnte entweder daran liegen, daß wir in unseren Erkenntnissen und methodischen Mitteln noch nicht so weit sind, um die gesetzeshaften Zusammenhänge zu erkennen, oder es ist grundsätzlich nicht möglich, Gesetze über alle Menschen zu allen Zeiten aufzustellen. Nach der letzteren Auffassung können sich Menschen auf Grund ihrer Sonderstellung in der Natur freiheitlich und reflektiert zu ihren Handlungen entscheiden und verhalten sich deshalb nicht nur in vorgezeichneten, für alle Menschen geltenden Bahnen. Betrachtet man sich nur das Faktum der Nicht-Existenz deterministischer Gesetzesaussagen in den Sozialwissenschaften, so hat dies weitreichende Konsequenzen für die Falsifikationstheorie Poppers.

Hypothesen mit einem Allgemeinheitsanspruch können zwar durch eine einzige empirische Falsifikation als falsch nachgewiesen werden, aber dies gilt nicht für probabilistische Hypothesen. In letzteren sind »Abweichungen« immer mit in der Hypothese enthalten (25 Prozent der Menschen zeigen laut Hypothese keine Verhaltensweisen, um ihre Freiheit nach einer Freiheitseinkengung wiederherzustellen); damit läßt sich aus einem Basissatz (Eine Person hat kein Freiheit wiederherstellendes Verhalten gezeigt) nicht die Falschheit eines probabilistischen Gesetzes nachweisen. Die logische Asymmetrie zwischen empirischer Verifikation und empirischer Falsifikation gilt nur für deterministische, nicht für statistische Hypothesen. Probabilistische Theorien werden

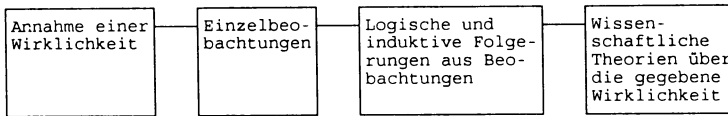
»falsifizierbar« gemacht, indem der Wissenschaftler zusätzlich Entscheidungen darüber trifft, wann der Widerspruch zwischen den gewonnenen Basissätzen (der statistisch interpretierten Ergebnislage) und der probabilistischen Theorie (eines als wahrscheinlich angenommenen Ergebnisbereiches) so groß ist, daß die Theorie als falsifiziert angesehen werden muß. In dieser notwendigen Vorgehensweise liegt auch die Erklärung dafür, daß nicht die Logik (z.B. mit ihren Ableitungsverfahren der Induktion und Deduktion), sondern die Statistik als methodisches Verfahren die Sozialwissenschaften dominiert.

Eine weitere Konsequenz des Fehlens deterministischer Gesetze in den Sozial-, Wirtschafts- und auch in einigen Bereichen der Naturwissenschaften besteht darin, daß bezogen auf Einzelereignisse bzw. Einzelpersonen präzise Erklärungen, Prognosen sowie genaue Wirkungsweisen von technologischen Anwendungen wissenschaftlich nicht zu leisten sind. Dies ist eine Folge der Wahrscheinlichkeitsaussagen in den probabilistischen Theorien, denn eine Wahrscheinlichkeitsaussage bezieht sich immer auf alle in den Aussagenbereich fallenden Ereignisse bzw. Sachverhalte, bei denen die aufgestellten Anfangsbedingungen zutreffen.

Fazit: Wichtig für den Erkenntnisanspruch von Wissenschaft war und ist die Einsicht innerhalb der Lehre des Kritischen Rationalismus, daß es keine absolut sichere Basis für Erkenntnis gibt (auch nicht durch empirische Basissätze, da diese durch konventionelle Einigungen belastet sind) – die Wahrheit von Sätzen und Theorien kann und darf deshalb nie in einem endgültigen Sinne behauptet und vertreten werden. Aufgrund des logisch unlösbaren Induktionsproblems ist letztlich keine Theorie beweisbar. Dies bedeutet aber, daß die Erklärungsansprüche von Theorien immer in einem vorläufigen Sinne verstanden werden sollten; Theorien gelten solange, bis sie eindeutig widerlegt werden oder bis Theorien entwickelt werden, die in ihrem Informationsgehalt, d.h. in der Reichweite ihrer Erklärungskraft, wesentlich über die alten Theorien hinausgehen.

Nach der bisher erfolgten Darstellung einiger Aspekte der Entwicklung der empirisch ausgerichteten Wissenschaftstheorien sollte jetzt auch verständlich sein, wieso die Position des Klassischen Empirismus/Positivismus bei dem heutigen wissenschaftstheoretischen Erkenntnisstand als »naiv« bezeichnet wird. Zur Verdeutlichung der Unterschiede sind in der Abbildung 1.2 die beiden grundlegenden wissenschaftstheoretischen Positionen gegenübergestellt. Bei einem Vergleich der beiden Schemata fällt auf, daß der Empirie im Forschungsablauf eine veränderte Rolle zugewiesen wird: Beobachtungen und Erfahrungen werden in der Wissenschaftsauffassung des Kritischen Rationalismus nicht mehr als der Ausgangspunkt wissenschaftlicher Forschung gesehen, sondern am Anfang der Forschung stehen immer Theorien. Die empirische Grundhaltung der »neueren Positionen« zeigt sich darin, daß der Erfahrungsbezug der wissenschaftlichen Theorien nachträglich durch die Bestätigung bzw. Verwerfung von aus den Theorien abgeleiteten Hypothesen hergestellt wird. Die Empirie hat die Funktion einer kontrollierenden Instanz gegenüber den

Das naive Wissenschaftsschemata:



Das Wissenschaftsschemata des Kritischen Rationalismus:

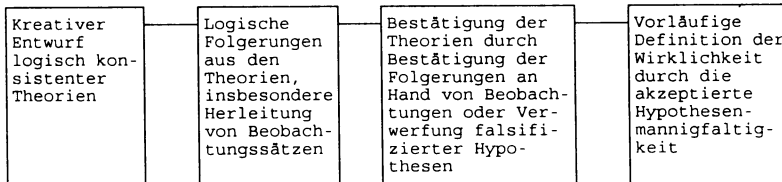


Abb. 1.2: Grundlegende Wissenschaftsschemata

spekulativen Theorieentwürfen zugewiesen bekommen.

In der Regel ist zur Erklärung eines Phänomens ein einzelnes, noch so allgemeines Gesetz nicht ausreichend. Üblicherweise gibt es für bestimmte Problembereiche (mehr oder weniger) bewährte *Systeme* von aufeinander bezogenen Erklärungen. Solche Erklärungssysteme nennt man *Theorien*. Unter Theorie wird im folgenden ein System logisch widerspruchsfreier Aussagen (Sätze, Hypothesen) über den jeweiligen Untersuchungsgegenstand mit den zugehörigen Definitionen der verwendeten Begriffe verstanden. In der Regel wird eine hierarchische Verknüpfung der Aussagen angestrebt: spezielle Aussagen werden in allgemeine Aussagen eingefügt bzw. sind aus ihnen abzuleiten. Einzelgefüge von Aussagen können dergestalt als *spezielle Theorien* bezeichnet werden, die Teiltheorien von *allgemeinen Theorien* darstellen.

»Alle Aussagen wissenschaftlicher Theorien stellen die Ergebnisse einer mehr oder minder weitgehenden Emanzipation von den Aussagen des Common Sense dar. Deren – oft logisch inkonsistentes – Gefüge wird als Menge von *Ethnotheorien* bezeichnet. Ihr steht die Menge der wissenschaftlichen Theorien gegenüber. Da Wissenschaftler das Alltagswissen ihrer Gesellschaft und Kultur teilen, bilden die Ethnotheorien die persönlichen und kulturellen Kontexte aller wissenschaftlichen Theorien und stellen gewissermaßen eine *kulturelle Matrix* dar, innerhalb welcher die wissenschaftlichen Theorien an bestimmten Orten verankert sind. Ethnotheoretisch selbstverständliche Grundperspektiven werden meist in ähnlicher Selbstverständlichkeit bezogen und so zum Ausgangspunkt aller wissenschaftlichen Begriffsbildung, Formulierung von Aussagen und Erstellung von Theorien. Auf diese Weise entsteht ein *Paradigma*, das sich dann in perspektivisch recht konsistenten Begriffen, Aussagen und Theorien konkretisiert« (Patzelt, W., 1986: Sozialwissenschaftliche Forschungslogik, München/Wien, S. 212f). Idealtypisch läßt sich die Struktur einer Theorie in einem Schema darstellen, wie es die Abbildung 1.3 zeigt.

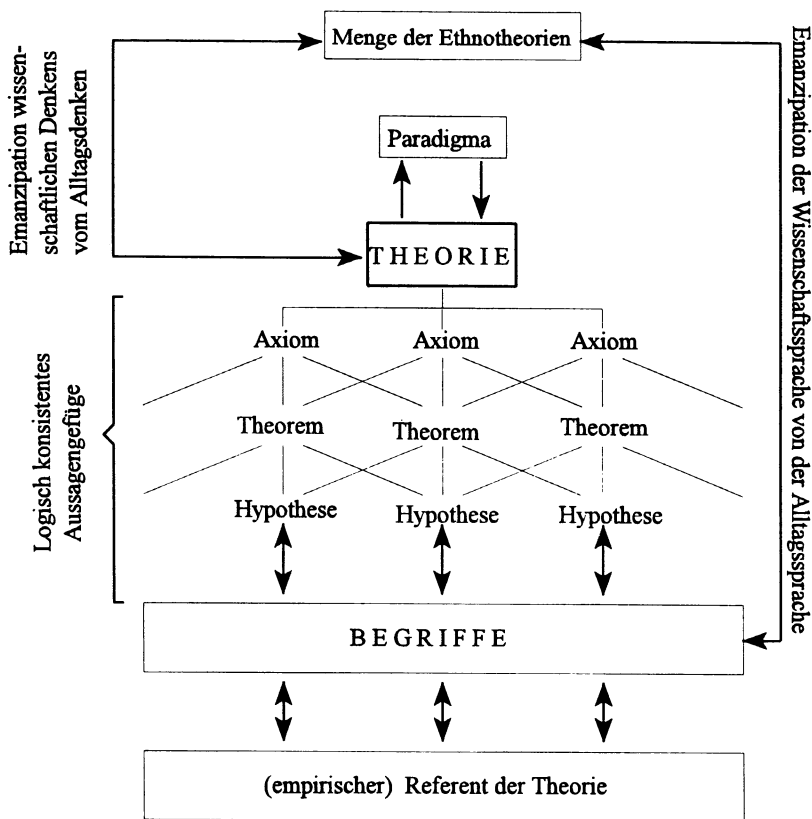


Abb. 1.3: Die Struktur einer Theorie

Quelle: Patzelt, W., 1986: Sozialwissenschaftliche Forschungslogik, München/Wien, S. 213.

Die *Axiome* (Postulate, Prämissen) sind grundlegende Sätze oder Aussagen, die gemeinsam den Informationsgehalt aller Aussagen einer Theorie in der Weise verdichten, daß sich alle weiteren Aussagen der Theorie logisch konsistent aus ihnen ableiten lassen. Für sie gelten folgende Regeln:

- Kein Axiom darf sich aus einem anderen ableiten lassen (jedes Axiom muß logisch unabhängig sein);
- eine Theorie muß genau so viele Axiome besitzen (im Grenzfall: eines), wie hinreichend und notwendig sind, um alle anderen Aussagen der Theorie logisch konsistent aus ihnen ableiten zu können.

Den Ausgangsbestand einer Theorie so zu ordnen, daß alle Einzelaussagen aus Axiomen dieser Art abzuleiten sind, heißt *Axiomatisierung* der fraglichen Theorie.

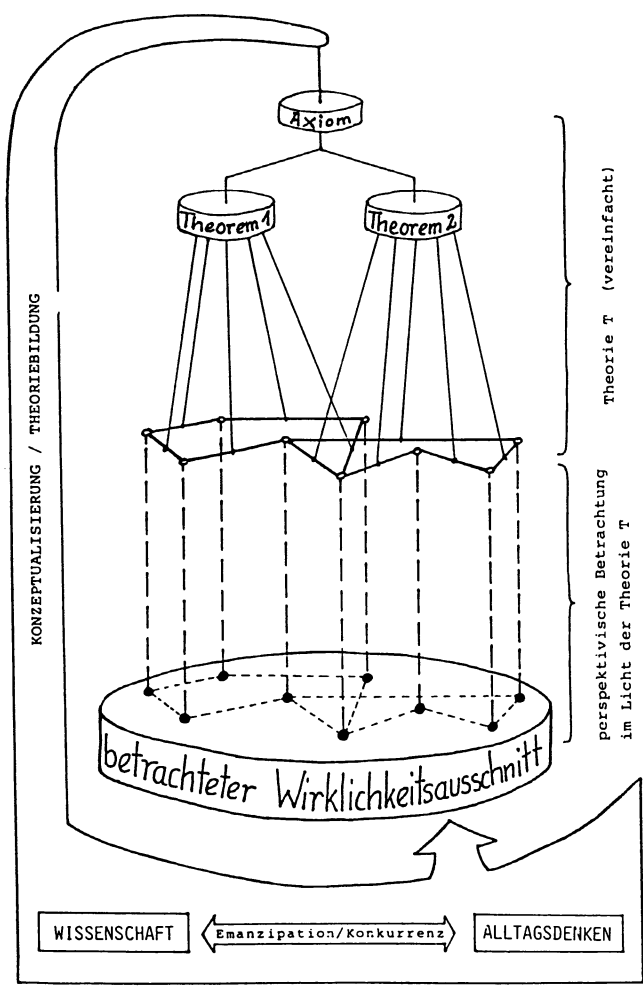
Mit Hilfe von logischen Transformations- und Ableitungsregeln können dann weitere Aussagen, die sog. *Theoreme* abgeleitet werden. Sie sind Aussagen, die den Informationsgehalt der hierarchisch unter ihnen stehenden Aussagen verdichten.

Bisweilen ist es möglich, den gesamten Informationsgehalt einer Theorie in sehr wenigen, hochabstrakten Axiomen oder Theoremen zu konzentrieren. Diese bilden dann den Kern der Theorie. Dieser *Theoriekern* selbst ist für die Bearbeitung konkreter Forschungsaufgaben eher nutzlos. Doch aus ihm läßt sich in jeder gewünschten Ausführlichkeit eine Vielzahl von Theoremen mit (forschungs-) gegenstandsbezogenen Aussagen und Hypothesen ableiten.

Die Aussagen in einem solchen theoretischen System können verschiedener Art sein; insbesondere betreffen sie unterschiedliche Abstraktionsstufen. Die Ableitung konkreter Sätze aus sehr allgemeinen Aussagen ist oft nötig, um allgemeine Aussagen empirisch überprüfbar zu machen. Eine empirische Referenz und damit einen Informationsgehalt bezüglich einzelner Wirklichkeitsausschnitte gewinnt der Theoriekern erst durch *gegenstandsspezifische Interpretation*, bei welcher aus dem Theoriekern die gegenstandsbezogenen Theorien abgeleitet werden, die ihrerseits die gesuchten, empirisch überprüfbaren Hypothesen enthalten.

Empirische Theorien lassen sich wie eine Netzkonstruktion denken, die über Wirklichkeitsausschnitten schwebt und von einem Mast gehalten wird, der die in Form bewußter Theoriebildung geleistete Emanzipation des wissenschaftlichen Denkens vom Alltagsdenken versinnbildlicht (s. Abbildung 1.4). Schon die Begriffe, mittels welcher empirische Theorien die »außen« bestehende Wirklichkeit erfassen, sind perspektivisch; ebenso verhält es sich mit den aus ihnen gebildeten Aussagen. Axiome und Theoreme verdichten die für eine bestimmte Theorie kennzeichnende Perspektive zu knappen Aussagen, aus denen sich (möglichst viele) Einzelaussagen mit hohem Informationsgehalt ableiten lassen. Innerhalb von Theorien werden theoretische (d.h. abstrakte) Begriffe auf *Beobachtungsbegriffe* (d.h. konkrete Begriffe) bezogen; durch Überprüfung des Wahrheitsgehalts der aus den Beobachtungsbegriffen gebildeten konkreten Aussagen kann man auf den Wahrheitsgehalt der aus den theoretischen Begriffen gebildeten abstrakten Aussagen schließen.

»Die gesamte 'Netzkonstruktion' der Theorie T bildet den Kontext jedes einzelnen Begriffs und jeder einzelnen Aussage, die zu dieser Theorie gehört. Im Kontext dieser Gesamtheorie werden auch alle einzelnen Beobachtungen bezüglich der empirischen Referenten der Begriffe und Aussagen von T gedeutet. In der Durchführung und Aufzeichnung derartiger Beobachtungen (Einholung von Informationen, Datenerhebung) besteht die empirische Forschung, die von T angeleitet wird, und deren Ergebnisse sowohl zur Überprüfung und Korrektur von T benutzt wird« (Patzelt, a.a.O., S. 218).



- Begriffe
- Hypothesen
- logisch konsistente Ableitungen von Aussagen aus anderen Aussagen
- Weg vom Begriff zu seinem empirischen Referenten
- perspektivisch betrachtete Sachverhalte im empirischen Referenten
- perspektivisch betrachtete Zusammenhänge im empirischen Referenten
- Annäherungen des Alltagsdenkens an den betrachteten Wirklichkeitsausschnitt

Abb. 1.4: Empirische Theorie und Wirklichkeit

Quelle: Patzelt, W., 1986: Sozialwissenschaftliche Forschungslogik, München/Wien, S. 219.

Empirische Theorien gliedert man zum einen nach den Gegenstandsbereichen, auf die sie sich beziehen. Dann unterscheidet man etwa soziologische, politikwissenschaftliche, wirtschaftswissenschaftliche usw. Theorien im allgemeinen oder nach spezifischen Wirklichkeitsausschnitten, etwa Rollentheorien, Parteientheorien, Konjunkturtheorien.

Andererseits lassen sich empirische Theorien danach einteilen, mit welcher Präzision sie einen wie umfangreichen Gegenstandsbereich zum empirischen Referenten haben. Folgende Arten empirischer Theorien sind unter diesem Gesichtspunkt zu unterscheiden (Patzelt, a.a.O., S. 220f):

- *Gegenstandsspezifische Theorien*: Sie decken mit großer Präzision einen eng umgrenzten Gegenstandsbereich ab (Beispiel: Theorien des Wahlverhaltens). Solche gegenstandsspezifischen Theorien, die vor allem Beobachtungsbegriffe enthalten und in erster Linie eine nach klaren Gesichtspunkten gegliederte Beschreibung eines Gegenstandsbereichs leisten, heißen *deskriptive Theorien*. Gegenstandsnahe Theorien, die in großem Umfang theoretische Begriffe enthalten und in erster Linie die Erfassung von erklärenden *Wenn/Dann*-Zusammenhängen ihres Gegenstandsbereichs anstreben, nennt man analytische Theorien².
- *Theorien mittlerer Reichweite* (*»middle range Theories«*): Sie decken größere Wirklichkeitsausschnitte ab als gegenstandsbezogene Theorien. Sie sind in allen Teilen bis hin zu konkreten Aussagen ausgearbeitet und lassen in hinreichender Weise eine Ableitung von empirischen Aussagen zu (Beispiel: Theorien des parlamentarischen Regierungssystems). Solche Theorien sind den weitausgreifenden Vernetzungen sozialer Wirklichkeit oft angemessener als engumgrenzte gegenstandsspezifische Theorien, welche größere Zusammenhänge in bisweilen erkenntnisverstellender Weise zertrennen müssen. Aufgrund des höheren Abstraktionsgrades sind Theorien mittlerer Reichweite eher analytisch als deskriptiv.
- *Allgemeine Theorien*: Sie decken sehr große Wirklichkeitsausschnitte ab, indem sie im Rahmen abstrakter Taxonomien eine Vielzahl von Aussagen auf hohem Abstraktionsniveau formulieren. Sie sind daher stets analytische Theorien. Allgemeine Theorien sind sehr hilfreich, um Wissensbestände zu integrieren und um weitverzweigte Gegenstandsbereiche überschauen zu können.

Der Wahrheitsgehalt einer Theorie wird durch die empirische Prüfung der aus ihr abgeleiteten Aussagen festgestellt. Dazu werden Informationen (*Daten*) über die Beschaffenheit jener Wirklichkeitsausschnitte erhoben, auf die sich die Behauptungen der zu prüfenden Aussagen beziehen. Um Aussagen überhaupt empirisch prüfen zu können, müssen sie Begriffe enthalten, die sich auf Beobachtbares beziehen (*Beobachtungsbegriffe*)³. Empirische Forschung setzt jene

² Der Begriff »analytisch« darf in diesem Zusammenhang nicht im Sinn »analytischer Aussagen« verstanden werden.

³ Der Begriff des »Beobachtens« soll im folgenden allgemein alle Arten der Feststel-

Leistung theoretischer Forschung voraus, die in der Definition präziser Beobachtungsbegriffe und in der Formulierung klarer Forschungsfragen besteht.

Im Rahmen einer erklärenden Sozialforschung kann empirische Forschung als Teil der *Theoriebegründung* gesehen werden. Empirische Forschung wird demnach hier in der Version des »Kritischen Rationalismus« als Überprüfungsinstrument theoretischer Aussagen dargestellt. Nach dieser Definition von »Wissenschaft« müssen die wissenschaftlichen Sätze objektive, d.h. empirisch überprüfbare Aussagen beinhalten, die stets den Charakter von Hypothesen haben, deren Informationsgehalt vom Grad ihrer Prüfbarkeit abhängig ist.

Eine *Hypothese* ist zunächst nicht mehr als eine Vermutung über einen Tatbestand. Forschungshypothesen können als die noch unbestätigte Form theoretischer Sätze aufgefaßt werden. Aus der Formulierung von Hypothesen müssen sich empirische Konsequenzen ableiten lassen, zu deren Nachweis oder Widerlegung die Methoden der empirischen Forschung notwendig sind. Zu einer Theorie gehört nun ein ganzes Hypothesenbündel, bestehend aus einem System von Einzelhypothesen über einen Gegenstandsbereich, das logisch in sich widerspruchsfrei ist. Das heißt, die einzelnen Sätze oder Hypothesen müssen sich auf den gleichen Gegenstandsbereich beziehen, und sie dürfen sich nicht logisch ganz oder teilweise gegenseitig ausschließen. Aus der Betonung explizit theorie-/hypothesengeleiteter Forschung folgt die Einbindung empirischer Forschung in einen deduktiven Ansatz: Die Formulierung von Hypothesen muß am Beginn der Forschungstätigkeit stehen. Eine Theorie/ein Hypothesenbündel kann nur an der Erfahrung scheitern, wenn sie schon vorher explizit gemacht ist.

Gültige und erfolgreiche Theorien entstehen meist erst nach langen Entwicklungen des *Versuchs und Irrtums*. Die in der analytisch-nomologischen Wissenschaftstheorie postulierte Wechselbeziehung Theorie-Empirie-Theorie erfolgt in einem schrittweisen Herantasten an »wahres« Wissen, das wiederum dokumentiert ist als ein logisch widerspruchsfreies System von empirisch bewährten Hypothesen. Dabei ist das sog. *Wahrheitskriterium* einzig und allein die Konfrontation mit erfahrbarer Realität bzw. erfahrbaren Realitätsausschnitten. Theorien müssen empirisch widerlegbar sein und sich dann unter unterschiedlichen Bedingungen bewähren. Theorieentwicklung kann so gesehen werden als eine Art von Evolution in einer unfreundlichen Umwelt angesehen werden (Popper, K.R., 1973: Objektive Erkenntnis. Ein evolutionärer Entwurf. Hamburg, S. 268): Jede Theorie beginnt mit einem Problem, das es zu lösen gilt P(1). Darauf erfolgen – über Versuch und Irrtum – vorläufige Lösungen dieses Problems in Form erster Erklärungen, die aber noch mit Fehlern behaftet sind. Empirische Tests und daran anschließende Veränderungen der Theorie führen – nach und nach – zur Beseitigung der Fehler. Gleichzeitig werden mit der Lösung des alten Problems ganz neue Fragen sichtbar. Gerade erst die

lung von Sachverhalten umfassen; er ist somit nicht auf die Methode der »Beobachtung« beschränkt.

Fehlerbeseitigung und die Lösung eines ersten Problems führt damit zur Formulierung eines neuen Problems P(2), das vorher nicht gesehen werden konnte. Und der ganze Prozeß beginnt von Neuem. Theorieentwicklung ist damit ein nie abzuschließender Prozeß der Lösung und der Neuformulierung von Problemen. Die wichtigste Rolle spielt dabei zum einen die logische Konsistenz (Widerspruchsfreiheit ihrer Aussagen) und zum anderen ihr empirischer Gehalt (die empirische Bewährung von Theorien in Untersuchungen).

Theoretische Kenntnisse oder die Hypothesen über bestimmte Sachverhalte und Beziehungen strukturieren erst den Gegenstandsbereich, lenken die Aufmerksamkeit auf bestimmte, als relevant angesehene Aspekte oder Dimensionen. Damit ist die wichtige Frage der gezielten Selektion von Merkmalen des Untersuchungsgegenstandes, die beobachtet werden sollen, angesprochen: Welche Merkmale sollen als besonders bedeutsam behandelt und somit erhoben werden? Erst wenn eine gezielte Auswahl vorgenommen wurde, können systematische Beobachtungen erfolgen und Daten erhoben werden. Nur eine systematische Beobachtung eines Objektbereichs kann gewährleisten, »daß regelhafte Beziehungen seiner Struktur oder Veränderung der Struktur ... erkannt werden. Eine Beobachtung ist prinzipiell theoriegeleitet; d.h. unabhängig von impliziten oder expliziten theoretischen Vorurteilen, ideellen Vorwagnahmen oder Vorentscheidungen ... kann keine Erfahrung gemacht werden« (Hülst, D., 1975: Erfahrung – Gültigkeit – Erkenntnis. Zum Verhältnis von soziologischer Empirie und Theorie, Frankfurt a.M., S. 14)⁴.

Ist eine Theorie/Hypothese empirisch zu testen, dann gibt sie bis zu einem gewissen Grade die Auswahl der zu beobachtenden Merkmale vor. Wir benötigen ein *deskriptives Schema*⁵, eine Begriffsanordnung, welche zu den Phänomenen und den Aspekten hinführt, denen die besondere Aufmerksamkeit gilt. Die Erstellung einer solchen Liste relevanter Eigenschaften, eines solchen deskriptiven Schemas, setzt wiederum zumindest implizite theoretische Überlegungen über die Eigenschaften des Gegenstandsbereiches und die Beziehungen zwischen diesen Eigenschaften voraus. An der Entwicklung dieser Liste objektiver Merkmale wird auch die wechselseitige Abhängigkeit von Theorie und Empirie deutlich: Je besser die theoretischen Kenntnisse, um so brauchbarer wird das deskriptive Schema, das die Erhebung lenkt. Je besser das deskriptive Schema, um so theoretisch relevanter werden die erhobenen Daten, um so besser sind die Voraussetzungen für die Fortentwicklung der Theorie. Die theoretischen Kenntnisse oder – wo solche nicht in gesicherter Weise vorhanden sind – Vermutungen, Hypothesen über bestimmte Sachverhalte und Beziehungen strukturieren erst den Gegenstandsbereich, lenken die Aufmerksamkeit auf bestimmte, als relevant angesehene Aspekte.

⁴ »Beobachtung« ist in diesem Zusammenhang in sehr weitgefaßtem Sinn zu verstehen als »jede Form sinnlicher Wahrnehmung«.

⁵ Vgl. Zetterberg, H., 1973: Theorie, Forschung und Praxis der Soziologie, in: König, R. (Hrsg.), 1973: Handbuch der empirischen Sozialforschung, Bd. 1, 3. A., Stuttgart, S. 118.

Aus diesen wenigen Überlegungen heraus wird bereits deutlich, wie sehr Hypothesenbildung bzw. Theoriekonstruktion, Datenerhebung und Auswertung (d.h. Anwendung von Erhebungstechniken und statistischen Modellen) miteinander verwoben und aufeinander angewiesen sind. Der Forscher wird bei den theoretischen Formulierungen bereits die Methoden der Datenerhebung und die möglichen statistischen Auswertungsmodelle mit bedenken müssen, ebenso wie er später bei der Berechnung statistischer Kennziffern ohne Rückbezug zu den theoretischen Überlegungen nur bedeutungslose Zahlen erhalten würde.

Zusammengefaßt: »Theorien sind für den Erfahrungswissenschaftler die wesentlichen 'Werkzeuge', die ihm seinen Zugang zur Realität ermöglichen:

- 1) Die Theorie liefert die grundlegende Orientierung; sie definiert den Objektbereich und sie legt fest, welche Aspekte der Realität zum Gegenstand der Forschungstätigkeit gemacht werden sollen.
- 2) Sie stellt das begriffliche Bezugssystem zur Verfügung; sie erlaubt, die als relevant definierten Aspekte (Dimensionen) des Objektbereichs systematisch darzustellen, zu klassifizieren und Beziehungen zu postulieren.
- 3) In Theorien werden empirisch ermittelte Fakten zu Generalisierungen bzw. zu Systemen solcher Generalisierungen zusammengefaßt. Je nach dem Grad ihrer Allgemeinheit kann man zwischen Ad-hoc-Theorien, Theorien mittlerer Reichweite und solchen höherer Komplexität (»allgemeinen« Theorien) unterscheiden.
- 4) Eine Theorie ermöglicht die Vorhersage zukünftiger Ereignisse. Wenn eine Zusammenfassung von Beobachtungen in Form von Verallgemeinerungen erstellt worden ist, erwartet man, daß dieselben Strukturen und Beziehungen auch dort gefunden werden, wo noch keine Erfahrungen oder Beobachtungen gemacht werden konnten, und man erwartet die Gültigkeit dieser Beziehungen auch in der Zukunft.
- 5) Schließlich gibt sie Hinweise auf vorhandene Wissenslücken«. (Kromrey, H., 1991: Empirische Sozialforschung, 5., überarb. u. erw. A., Opladen, S. 46).

Zum richtigen Verständnis des Erkenntnisprozesses im weiteren und des empirischen Forschungsprozesses im engeren Sinne ist es wichtig, diesen Prozeß als einen interdependenten Vorgang zu begreifen (s. Abbildung 1.5). In den Sozialwissenschaften geht es darum, Informationen systematisch zu erheben, zu organisieren und zu analysieren, um gesellschaftliche Sachverhalte adäquat beschreiben und erklären zu können. Ausgangspunkt sind dabei zwei Pole, die in der Abbildung 1.5 durch die beiden großen Rechtecke auf der linken und rechten Seite dargestellt sind. Auf der einen Seite steht die soziale Realität, die erfahrungswissenschaftlich begriffen werden soll, auf der anderen Seite stehen die Vorstellungen von der sozialen Realität oder eine sozialwissenschaftliche Theorie. Den Teilbereich aus diesen allgemeinen Vorstellungen, der den eigentlichen Erkenntnisgegenstand einer konkreten Untersuchung betrifft, bezeichnet

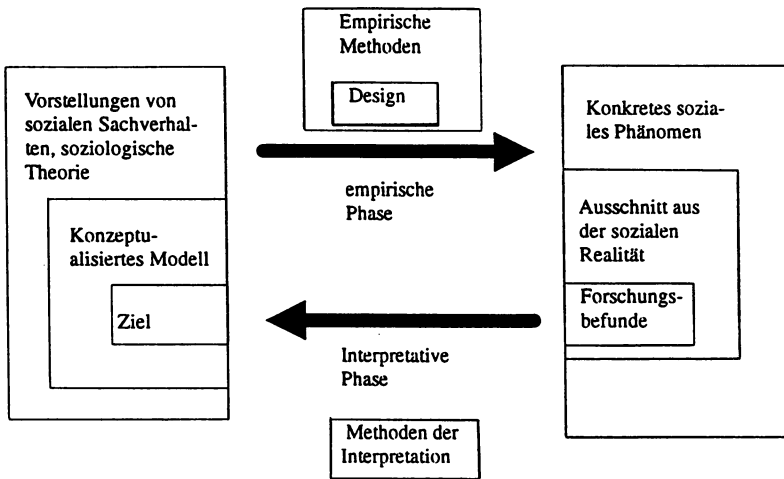


Abb. 1.5: Verhältnis Theorie und Empirie

Quelle: Riley, M.W., 1963: Sociological research. A case approach, New York, S. 4.
(Hier in leicht abgewandelter Form.)

man als *Haupttheorie*. Diese theoretischen Vorstellungen (repräsentiert als Hypothesenbündel) gelangen in Verbindung zur sozialen Realität durch das *Erfahrungswissen*, repräsentiert durch das unmittelbare Ergebnis einer Untersuchung. Es handelt sich um Aussagen über konkrete Tatbestände, die als Protokollaussagen, singuläre Sätze oder Daten bezeichnet werden. Sie sind auf der rechten Seite der Abbildung als Forschungsbefunde repräsentiert. Der empirische Forschungsprozeß soll nun die Vorstellungen von der sozialen Realität (repräsentiert als Theorie, Theorieausschnitt oder Hypothesenbündel) den sozialen Sachverhalten annähern: durch Präzisierung der Aussagen, Ausfüllen von Lücken, Auflösung verzerrter oder falscher Aussagen sowie Absicherung zutreffender Vorstellungen. Dazu müssen die Forschungsbefunde (repräsentiert in Form von 'statistischen' Ergebnissen) in ein sozialwissenschaftliches 'Ergebnis' rückübersetzt werden – dieser Vorgang heißt Interpretation. Man kann diese Vorgehensweise auch als Problemlösungsprozeß verstehen, bei dem es ganz allgemein um eine sukzessive Annäherung der theoretischen Vorstellungen an den jeweils betrachteten Realitätsausschnitt geht.

Ein konkretes wissenschaftliches Problem ergibt sich z.B. zum einen aus der Konkurrenz verschiedener Theorien bei der Beschreibung oder Erklärung desselben Gegenstandsbereiches; zum anderen kann unser Erfahrungswissen mit dem theoretischen Wissen konfligieren, sei es als Widerspruch zu theoretischen Aussagen, sei es als Erklärungslücke.

Der aktive Erkenntnisprozeß geht dabei von dem bereits bestehenden Vorverständnis aus, das Erfahrung strukturiert. Die so strukturierte Erfahrung schlägt sich in Daten nieder, die wiederum mit Hilfe des Überprüfungsinstruments »Empirischer Forschungsmethoden« zur Erweiterung oder Korrektur des Vorverständnisses beitragen.

Die Daten selbst sind allerdings nicht immer unproblematisch in dem Sinne, als die Anwendung von Verfahrensweisen der empirischen Sozialforschung (Datenerhebungstechniken, Stichprobenverfahren, Datenanalysetechniken etc.) zum einen mit Voraussetzungen verbunden und zum anderen mit vielfachen Problemen und Grenzen behaftet sind. Die Erfahrungsdaten sind zwar die entscheidende Grundlage für die Überprüfung theoretischer Aussagen, sie sind aber selbst durch methodologische Strukturen geprägt und bis zu einem gewissen Grad begrenzt oder verzerrt. Der Erkenntnis- und Forschungsprozeß besitzt also kein absolutes Fundament, sondern muß sich auf einer mit Fehlern behafteten Datengrundlage durch beständiges Überprüfen beider Seiten vorwärts entwickeln.

Schlußbemerkung:

Die hier vorgestellte Forschungslogik vertritt – im Anschluß an die Wissenschaftstheorie des »Kritischen Rationalismus« – einen methodologischen Rigorismus, wie er in der methodologischen Praxis der historischen Sozialforschung nur in Ausnahmefällen vollständig und konsequent einzulösen sein wird. Diese Forschungslogik ist als eine normative Anleitung zu verstehen, die eine systematische Planung und eine intersubjektiv gültige bzw. überprüfbare Durchführung von empirischen Forschungen gewährleistet. Abweichungen von dieser »Norm« in der Forschungspraxis sind oft unvermeidlich; die durch die Forschungslogik vorgegebene Transparenz der Vorgehensweise erlaubt jedoch, diese »Abweichungen« überhaupt festzustellen und dann zu entscheiden, inwieweit sie noch tolerierbar bzw. inwieweit Änderungen in der Vorgehensweise notwendig sind.

2. Arten sozialwissenschaftlicher Forschungsdesigns

Sozialwissenschaftliche Forschung läßt sich zu ganz verschiedenen Zwecken und auf unterschiedlichste Art durchführen. Um einen Überblick zu geben, sollen im folgenden die zehn von Patzelt entwickelten Gesichtspunkte zur Klassifikation sozialwissenschaftlicher Forschungsarbeiten wiedergegeben werden (Patzelt, W.J., 1986: Sozialwissenschaftliche Forschungslogik, München/Wien, S. 299ff). »Das Klassifikationsschema führt zu einem zehn-dimensionalen Merkmalsraum, innerhalb dem für jede einzelne sozialwissenschaftliche Studie die für sie nützliche Merkmalskombination zu wählen ist. Damit ist jedoch nicht an eine Schematisierung der Forschungspraxis gedacht, denn jede Studie wird in hohem Maße individuell sein. Eine Klassifikation bietet vielmehr die

Möglichkeit einer genauen Verortung des zugrundeliegenden Untersuchungsdesigns.«

- I. Grundlegend ist zunächst zwischen *theoretischer* und *empirischer Forschung* im engeren Sinn zu unterscheiden.
- II. Ferner ist zwischen *normativer Forschung* und *empirischer Forschung* im weiteren Sinn zu unterscheiden.
- III. Forschungsarten lassen sich auch nach ihrer besonderen *Aufgabenstellung* gliedern. Unter diesem Gesichtspunkt sind hervorzuheben:
 - a) *Deskriptive Studien* bezwecken die Erarbeitung von Tatsachen- und Zusammenhangswissen; sie machen Aussagen über die Beschaffenheit sozialer Wirklichkeit.
 - b) *Analytische Studien* streben Erklärungsaussagen an, mit deren Hilfe Retrodiktionen oder Prognosen erstellt werden können. Sie nehmen meist generalisierenden Charakter an.
 - c) In *theorieprüfenden Studien* werden Theorien auf ihren logischen und empirischen Wahrheitsgehalt überprüft. Beziehen sich die geprüften Theorien auf grundlegende Sachverhalte sozialer Wirklichkeit, so arbeitet man im Bereich der sozialwissenschaftlichen Grundlagenforschung.
 - d) *Anwendungsorientierte Studien* beziehen anhand bewährter normativer und empirischer Theorien das verfügbare Tatsachen-, Zusammenhangs- und Erklärungswissen auf einzelne Probleme und entwickeln dann Handlungsanweisungen.
- IV. Forschungsarbeiten werden auch danach unterschieden, *wieviele Untersuchungseinheiten* sie auf welche Weise betrachten.
 - a) Im Grenzfall liegt eine *Einzelfallstudie* vor. Hier liegt der Versuch vor, den Wahrheitsgehalt von Erklärungen und Theorien bezüglich eines Einzelfalls durch Verifikation oder Falsifikation festzustellen. Am geeignetsten sind Einzelfallstudien für die deskriptive Forschung.
 - b) Den Gegentyp stellen *vergleichende Studien* dar, in denen simultan mehrere Untersuchungseinheiten entlang forschungsrelevanter Merkmale verglichen werden. Vier Untergattungen sind hervorzuheben:
 - *Querschnittsstudie*: Hier werden zu einem bestimmten Zeitpunkt verschiedene Untersuchungseinheiten anhand gleicher Variablen betrachtet.
 - *Historisch vergleichende Studien*: Bei ihnen werden anhand gleicher Variablen verschiedene Untersuchungseinheiten aus unterschiedlichen Zeiten betrachtet. Dieses Untersuchungsdesign ermöglicht Aussagen über historische Besonderheiten. Bei entsprechend gewählten Untersuchungseinheiten lassen sich auch Aussagen über geschichtliche Entwicklungen prüfen.

- *Regional bzw. national vergleichende Studien*: Bei ihnen werden anhand gleicher Variablen unterschiedliche Untersuchungseinheiten aus verschiedenen Regionen oder Staaten betrachtet. Ziel ist die Analyse von regionalen oder nationalen Gemeinsamkeiten und Besonderheiten.
- *Sektoral vergleichende Studien*: Bei ihnen werden anhand gleicher Variablen unterschiedliche Untersuchungseinheiten aus verschiedenen Sektoren desselben sozialen Systems betrachtet. Dieses Design ermöglicht die Ermittlung von besonderen Problemlagen wie auch die Gemeinsamkeiten einzelner Sektoren.

Vergleichende Studien lassen sich, je nach Forschungsziel, in beliebiger Komplexität nach mehreren Gesichtspunkten anlegen. Regional, national und sektoral vergleichende Studien werden zusammenfassend auch 'Querschnittsstudien' genannt und den anschließend zu behandelnden *Längsschnittstudien* gegenübergestellt. Historisch vergleichende Studien sind dann Längsschnittanalysen, wenn man die zu vergleichenden Untersuchungseinheiten und Zeitpunkte so wählt, daß konkrete Prozesse geschichtlichen Wandels erkennbar werden.

- V. Forschungsarten lassen sich wie folgt nach ihrer *zeitlichen Perspektive* gliedern:
- a) Werden vor allem *Zustände* (weniger die Wandlungsprozesse zwischen bestimmten Zuständen) untersucht, dann liegen synchrone Studien vor, die in den meisten Fällen Querschnittsuntersuchungen sind.
 - b) Werden vor allem *Wandlungsprozesse* untersucht, so liegen diachrone Studien ('Längsschnittuntersuchungen') vor. Sie lassen sich nach der Weise, wie der Wandel erfaßt wird, wie folgt unterscheiden:
 - Wandel kann durch die *Beschreibung von Prozessen* untersucht werden. Hierfür sind oft die statistischen Modelle der Zeitreihenanalyse anwendbar.
 - Wandel kann durch historisch vergleichende Studien, also durch den Vergleich *aufeinanderfolgender Zustände*, erfaßt werden. Sie nehmen oft eine der folgenden Formen an:
 - *Folgestudien* liegen dann vor, wenn Daten über dieselben Variablen zu verschiedenen Zeitpunkten an jeweils neuen Stichproben aus der Grundgesamtheit erhoben werden.
 - Eine *Panelstudie* liegt vor, wenn dieselben Untersuchungseinheiten zu verschiedenen Zeitpunkten entlang dem gleichen Variablen set erhoben werden. Bei Panelstudien verfügt man über Zeitreihendaten zu den einzelnen Untersuchungseinheiten.
 - Eine *Kohortenstudie* liegt vor, wenn Wandlungsprozesse in bezug auf voneinander unterscheidbaren Generationen betrachtet werden. Eine 'Generation' im sozialwissenschaftlichen Sinn ist eine Gruppe von Personen, die z.B. in ihrer Kindheit und Jugend durch gemeinsam

erlebte Ereignisse geprägt wurde, die in der Sozialisation älterer Menschen keine ähnliche Rolle spielten. Da angenommen wird, daß manche Wandlungsprozesse von Menschen je nach ihrer sozialisationistischen Prägung anders erfahren werden, verarbeitet und vor allem auch anders gestaltet werden, versucht man, diesen 'Generationeneffekt' dadurch zu berücksichtigen, daß man etwa innerhalb von Folgestudien einzelne Generationen getrennt untersucht. Alternativ können auch die Geburtsjahrgänge einer Querschnittsuntersuchung zu 'Generationen' zusammengefaßt werden. Solche getrennt untersuchten Generationen heißen 'Kohorten'.

- VI. Forschungsarten werden auch nach ihrem *Stellenwert innerhalb der längerfristigen Erforschung eines Gegenstandsbereichs* unterschieden:
- a) Wenn noch wenig Vorwissen verfügbar ist, also weder präzise Begriffe den Forschungsprozeß anleiten können noch der Ablauf des Forschungsprozesses genau zu planen ist, sondern es vielmehr auf erste Informationen über die Beschaffenheit eines bestimmten Gegenstandsbereichs ankommt, dann wird eine *explorative Studie* durchgeführt.
 - b) Wenn ein komplexer Gegenstandsbereich nur dem Ansatz nach, doch aus zeitlichen oder finanziellen Gründen (noch) nicht in der gewünschten oder sachlich notwendigen Vielschichtigkeit untersucht werden kann, führt man eine *Pilotstudie* ('Vorstudie') durch. Bei ihr werden zwar angemessene Erhebungsinstrumente entwickelt und verwendet, doch bei der Datenerhebung beschränkt man sich einerseits auf nur eine begrenzte Anzahl von Untersuchungseinheiten, und andererseits verzichtet man gegebenenfalls darauf, an diesen Untersuchungseinheiten sämtliche, theoretisch erforderlichen Variablen zu erheben. Die Ergebnisse einer Pilotstudie geben daher nur erste Hinweise auf die Beschaffenheit des untersuchten Gegenstandsbereichs, können aber keinesfalls als 'bewährte Forschungsergebnisse' gelten.
 - c) Sozialwissenschaftliche Erhebungsinstrumente erzeugen bei ihrer Anwendung nicht selten Probleme, an die man bei ihrer Entwicklung trotz aller Sorgfalt nicht dachte. Deshalb ist es ratsam, die Erhebungsinstrumente vor der eigentlichen Datenerhebung zu erproben. Eine zu diesem Zweck durchgeführte ('probeweise') Datenerhebung heißt *Pretest*. Dabei werden die Erhebungsinstrumente auf Anwendungsschwierigkeiten, nicht bedachte Lücken hinsichtlich der zu erhebenden Informationen etc. überprüft. Im Licht der bei einem Pretest erworbenen Erfahrungen nimmt man gegebenenfalls Veränderungen an den Erhebungsinstrumenten vor.
 - d) Als *Hauptuntersuchung* bezeichnet man eine empirische Studie, die auf der Grundlage ausreichender Explorationsforschung, Pretests oder gegebenenfalls erst im Anschluß an eine Pilotstudie durchgeführt wird.
 - e) *Replikationen* sind Wiederholungen von Hauptuntersuchungen. Sie können einerseits in Form von Folgestudien zur Erfassung von Wandel, an-

dererseits zur Überprüfung von bereits vorliegenden Forschungsergebnissen dienen.

VII. Empirische Studien lassen sich auch nach der *Anzahl der bei ihnen verwendeten Methoden* unterscheiden. Aufgrund der Komplexität sozialwissenschaftlicher Forschungsgegenstände sind oftmals *Mehr-Methoden-Untersuchungen* angemessener. Diese sind zwar zeit- und kostenaufwendig und sprengen oft jenen Rahmen, der den sozialwissenschaftlichen Forschungsmöglichkeiten im konkreten Fall gesetzt ist. Da jede Erhebungsmethode nur eine jeweils besondere Art von Daten zu erheben erlaubt, ergeben sich möglicherweise für eine konkrete Fragestellung Begrenzungen, die durch eine Methodenkombination abgeschwächt oder aufgehoben werden können.

VIII. Den Aufbau sozialer Wirklichkeit kann man sich auch nach verschiedenen *Untersuchungsebenen* gegliedert vorstellen. Folgende Ebenen lassen sich unterscheiden:

- handelnde Einzelpersonen ('Akteure');
- Handlungsgeflechte im Umkreis von ausgewählten Akteuren ('soziale Netzwerke');
- räumliche Einheiten, in denen Einzelpersonen oder Netzwerke handeln ('räumliche Aggregate');
- einzelne Organisationen oder Institutionen;
- Organisations- und Institutionengefüge innerhalb von umfassenden sozialen Systemen (z.B. innerhalb eines Staates);
- Einzelstaaten;
- multi-staatliche und inter- bzw. transnationale Handlungsgeflechte.

Manche Untersuchungsgegenstände der sozialwissenschaftlichen Forschung sind auf nur jeweils einer Untersuchungsebene angesiedelt; manche Fragestellungen umspannen aber auch mehrere Untersuchungsebenen. Dementsprechend lassen sich sozialwissenschaftliche Studien nach der Anzahl der in sie einbezogenen Analyseebenen als Ein- und Mehrebenen-Studien einteilen. Mehrebenen-Studien werden in der Regel auch Mehr-Methoden-Studien sein müssen.

IX. Je nach dem Forschungszweck werden die erforderlichen Daten in Form einer *Vollerhebung* an allen zu betrachtenden Untersuchungseinheiten erhoben (die Grundgesamtheit) oder lediglich an einer *Auswahl* ('Stichprobe' aus der Grundgesamtheit). Nach der Art der Auswahl unterscheidet man grob:

- willkürliche Auswahlen;
- bewußte Auswahlen;
- reine Zufallsauswahlen;
- Ersatzverfahren für reine Zufallsauswahlen.

Allein die Zufallsstichproben sind 'echte' *Repräsentativ-Stichproben*, die

einen (wahrscheinlichkeitstheoretisch) begründeten Repräsentations- oder Inklusionsschluß von der Stichprobe auf die Grundgesamtheit erlauben.

- X. Schließlich lassen sich empirische Studien auch nach ihrem *Empirie-Bezug* gliedern, d.h. nach ihrer Nähe zur tatsächlichen Beobachtung jener Wirklichkeitsausschnitte, über die sie Aussagen treffen.
- a) Am engsten ist der Empirie-Bezug bei Studien, denen eine konkrete Datenerhebung vorausging. Solche Studien nennt man *empirische Originalarbeiten*. Sie stellen die Hauptaufgabe empirischer Forschung dar. Der Wert einer empirischen Studie für eine zielsichere und unmißverständliche wissenschaftliche Kommunikation wird durch einen standardisierten Aufbau vergrößert. Ein Forschungsbericht sollte die folgenden, klar voneinander getrennten Abschnitte enthalten:
- Darlegung der Fragestellung, des Forschungskontextes und der gewählten theoretischen Perspektive;
 - Darstellung des Auswahlverfahrens, der verwendeten Methode sowie des datenanalytischen Vorgehens in allen für den Kommunikationspartner wichtigen Einzelheiten;
 - Dokumentation der Forschungsergebnisse in Tabellen und Schaubildern;
 - Diskussion und Interpretation der Ergebnisse, wobei die eingangs formulierten Forschungsfragen beantwortet und weiterführende Forschungsaufgaben genannt werden.
- b) Die empirischen Originalstudien werden für *empirische Studien im weiteren Sinn* sekundäranalytisch ausgewertet; es entstehen 'empirische Sekundärstudien'. In ihnen werden die Forschungsergebnisse, auf die man sich bezieht, nicht mehr in unmittelbar nachprüfbarer Weise dargelegt, sondern man zitiert diese Ergebnisse nur und verwendet sie – gegebenenfalls in Form von Tabellen oder Statistiken – an geeigneter Stelle im eigenen Gedankengang. Ein großer Teil der sozialwissenschaftlichen Literatur besteht aus derartigen empirischen Sekundärstudien.
- c) Unter *Sekundäranalysen* versteht man auch Forschungsarbeiten, die sich direkt auf die Datengrundlage einer empirischen (Haupt-) Studie stützen⁶. Der Forschungsprozeß schließt die aktive Auswertung eines Datensatzes aus einer Primärstudie ein, an der der Sekundär-Forscher selbst gar nicht beteiligt war. Folgende Vorteile können sich aus dieser Vorgehensweise ergeben:
- Verwendung der Primärdaten für Ausbildungszwecke (etwa in der Me-

⁶ Das Zentralarchiv für Empirische Sozialforschung (Universität zu Köln, Bachemerstr. 40, 50931 Köln), stellt archivierte Daten bereit, die für weitere Analysen erneut quantitativ ausgewertet werden können. Der Arbeitsbereich des Zentralarchivs erstreckt sich auf alle Fachgebiete, in denen Verfahren der empirischen Sozialforschung verwendet werden. Auskünfte über die Bestände gibt ein detaillierter Datenbestandskatalog.

thodenausbildung);

- bessere Ausschöpfung des Informationsgehaltes der Primärstudie;
- Beantwortung der Forschungsfrage mit Hilfe alternativer Datenanalyseverfahren;
- Verkettung mehrerer Datensätze mit gleichen Variablen um z.B. zeitliche oder regionale Vergleiche bezüglich einer bestimmten Fragestellung durchführen zu können.«

3. Der Forschungsprozeß: Ein zusammenfassender Überblick

Die Methoden der empirischen Forschung sind Verfahrensregeln formaler Art, die sicherstellen sollen, daß ihre Ergebnisse einen angebbaren Grad von Verbindlichkeit haben. Zu diesem Zweck empfiehlt sich ein bestimmtes Vorgehen bei der empirischen Untersuchung von Objektbereichen. Dieses Vorgehen bezieht sich auf Entscheidungen, die im *Forschungsprozeß* zu treffen sind. Bei der Durchführung eines empirischen, sozialwissenschaftlichen Forschungsprojektes lassen sich idealtypisch einige notwendige Entscheidungen angeben (s. Abbildung 3.1). Das Modell des Forschungsprozesses enthält zwei Ebenen, auf denen sich der Ablauf vollzieht. Eine Ebene beinhaltet die Reflexionen über Ablauf und Prinzipien des Forschungsprozesses selbst (Meta-Ebene) und die theoretischen Zusammenhänge, in die der Forschungsprozeß eingebettet ist. Die zweite Ebene betrifft die Fragen der konkreten Forschungsschritte.

Die dargestellte Abfolge bildet ein forschungslogisch und forschungspraktisch begründetes Raster, das allerdings im realen Forschungsprozeß nicht immer starr befolgt werden kann. Im realen Forschungsprozeß werden Überschneidungen die Regel sein. Die Entscheidungen sind vielfältig ineinander verzahnt. Häufig wird in einer fortgeschrittenen Forschungsphase die Rückkehr zu vorgeordneten Entscheidungspunkten erforderlich.

Jedes Forschungsvorhaben impliziert bis zum Abschluß eine Fülle von Entscheidungen über den Umgang mit anfallenden Problemen (Auswahl des Forschungsproblems, Wahl des theoretischen Bezugsrahmens, Wahl der Erhebungsinstrumente, Auswahl der Untersuchungsobjekte, Anwendung statistischer Verfahren usw.), die erhebliche Auswirkungen auf das Forschungsergebnis haben können und deshalb zur Sicherung der intersubjektiven Überprüfbarkeit sorgfältig zu dokumentieren sind (ausführlich s. Schrader, A., 1971: Einführung in die empirische Sozialforschung, Stuttgart/Berlin.). Im folgenden soll der Ablauf detaillierter dargestellt werden.

a) Die Wahl des Forschungsproblems

In der Regel beginnt ein Forschungsprojekt mit der Festlegung des Gegenstandes der Forschung, der Formulierung des Forschungsproblems. In dieser Phase

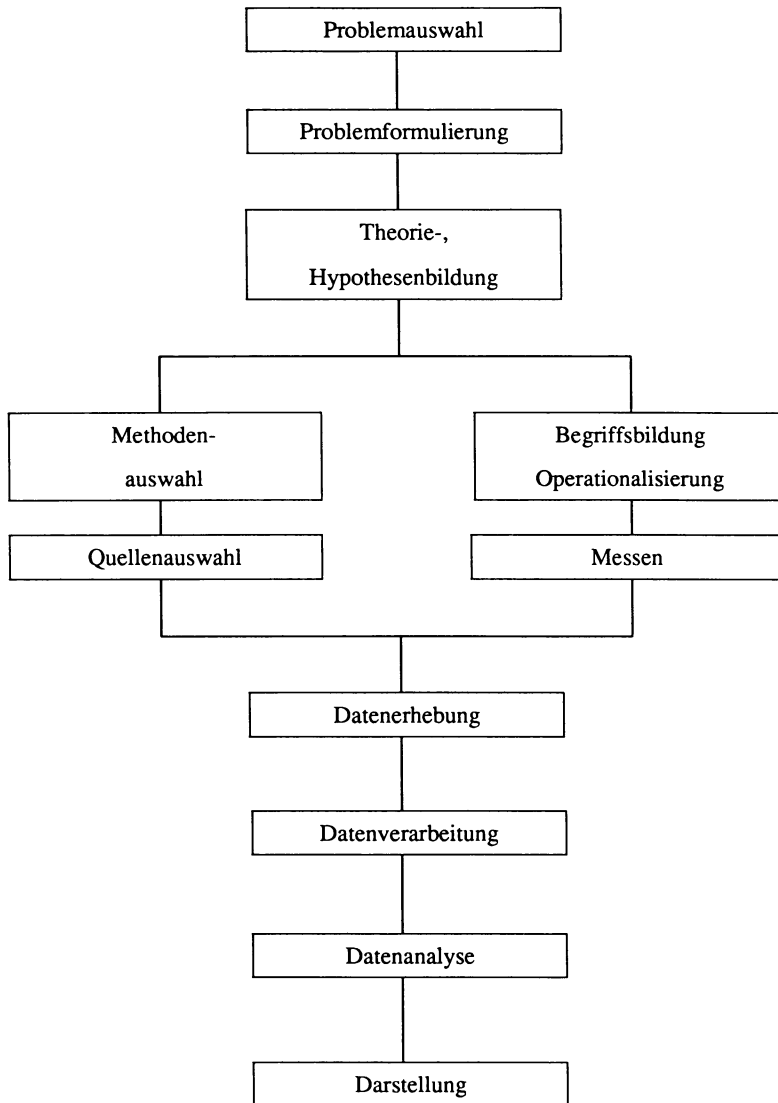


Abb. 3.1: Der Forschungsprozeß

sind insbesondere Fragen des *Entdeckungs-* und des *Verwertungszusammenhanges* von Bedeutung. Man kann hier danach unterscheiden, ob es sich um eine Auftragsforschung handelt (nach Vorgabe durch den Auftraggeber, z.B. von Landes- und Bundesministerien/-behörden, Verbänden, Gemeinden, Gewerkschaften und anderen Einrichtungen), oder ob es sich um ein von den Forschern selbst initiiertes Forschungsprojekt handelt. Viele Anlässe können einen Forschungsprozeß in Gang setzen: ein plötzlich auffallender Sachverhalt, ein Betroffenheit erzeugendes Problem, die Chance, Geld, Personal, Kontakte oder ein 'prominentes Thema' zu erlangen. Welche Fragen man aufgreift und in Forschungsleistungen umsetzt, hängt auch von wertgeleiteten persönlichen Entscheidungen des Forschers ab, die ihrerseits von zufälligen Möglichkeiten und rahmensetzenden Zwängen beeinflußt werden.

Die konkrete Wahl eines bestimmten Untersuchungsgegenstandes hängt im Fall einer Forschungstradition stark von der jeweils aktuellen Forschungssituation ab, wie sie sich in der Fachliteratur zeigt. Die zu bearbeitenden Fragestellungen können in vielerlei Hinsicht variieren:

- Eine empirische Untersuchung kann mit dem Ziel des Testens von wissenschaftlichen Theorien/Hypothesen oder zur empirischen Entscheidung über die Angemessenheit konkurrierender Theorien durchgeführt werden.
- Eine empirische Untersuchung kann auch dazu dienen, das wissenschaftliche Grundlagenwissen zu erhöhen.
- In einem relativ neuen Problemfeld, für das es bisher nur wenig gesichertes Wissen gibt, können z.B. erstmals Basisdaten beschafft werden. In diesem Fall liegt eine deskriptive Fragestellung vor.
- Das Erkenntnisinteresse kann aber auch darauf gerichtet sein, eine möglichst exakte Beschreibung eines komplexen Sachverhalts zu gewinnen, über den zwar schon ein hinreichendes (Rahmen-) Wissen existiert, über den jedoch für einen bestimmten Zeitpunkt/-raum und/oder einen bestimmten räumlichen Bereich präzisere Informationen benötigt werden.

Die Frage des *Verwertungszusammenhanges* steht für die Zwecke, für die die Ergebnisse verwendet werden sollen. Die erarbeiteten Ergebnisse lassen sich im Wissenschaftsbetrieb oder im Rahmen der ihn ermöglichenden gesellschaftlichen, politischen, wirtschaftlichen oft für weitere Zwecke nutzen und verwerten. Nach Maßgabe eines Auftrags oder der Einschätzung der praktischen Auswirkungen erzielter Forschungsergebnisse kann sich der Forscher für solche Anwendungsmöglichkeiten interessieren, wenn nicht sogar für ihre Verwertung (mit-) verantwortlich fühlen. Besteht man auf einer solchen *Verantwortlichkeit des Forschers für die Verwertung seiner Ergebnisse*, so ist auch die *Nutzungsphase* als Abschnitt des Gesamtprozesses der Forschung anzusehen.

b) Theorie-/Hypothesenbildung

Die Vorstudienphase setzt ein, sobald im Entdeckungszusammenhang ein konkretes Forschungsinteresse entstand. Um den Forschungsbedarf überhaupt erst einmal festzustellen, ist zunächst das bereits verfügbare Wissen eines spezifischen Gegenstandsbereichs zu sichten. Dies geschieht durch *Bibliographieren und Einlesen*, d.h. als Sichtung und Bewertung der zum Forschungsthema existierenden Fachliteratur, insbesondere in Form von Fachzeitschriftartikeln⁷: Bereits verfügbare Aussagen, Theorien oder Theorieausschnitte müssen überblickt, auf Plausibilität geprüft und für die konkrete Forschungsfrage vernetzt werden; bereits vorliegende empirische Untersuchungen müssen sekundäranalytisch ausgewertet werden. Insgesamt wird es zu einem immer besser werdenen Verständnis des zu erforschenden Gegenstandsbereichs kommen. Nicht selten findet eine Neuformulierung der ursprünglichen Forschungsfrage statt oder es fließen bisher unbeachtete Aspekte, neuartige Zusatzfragestellungen mit in die Studienplanung ein.

In der Phase der Theoriebildung greift der Forscher entweder auf bereits ausgearbeitete Theorien (oder Theorieausschnitte) zu einem bestimmten Gegenstandsbereich in der Literatur zurück, oder es muß erst eine neue Theorie zur Erklärung des ausgewählten Gegenstandsbereichs entwickelt werden. Falls nicht auf eine explizite Theorie zurückgegriffen werden kann, läßt sich oft eine Übertragung von theoretischen Aussagen aus verwandten Gegenstandsbereichen vornehmen. Scheidet diese Vorgehensweise aus, so bieten vielfach verschiedene (allgemeine) theoretische Ansätze hinreichende Ansatzpunkte für eine Theoriebildung zu einem spezifischen Gegenstandsbereich. Ziel ist es, auf den zu untersuchenden Gegenstandsbereich zugeschnittene, möglichst exakte und nachprüfbar Hypothesen zu entwickeln. Auch bei diesem Schritt wird man prüfen, ob es bereits Hypothesen aus anderen, vergleichbaren Untersuchungen gibt, die herangezogen werden können. Theorie- und Hypothesenbildung soll zu einer modellhaften Strukturierung des Objektbereichs beitragen: In gedanklichen Vorleistungen ist der Forschungsgegenstand in seine vielfältigen Facetten zu zerlegen und zu ordnen, so daß daraus ein problemadäquates Forschungsdesign entwickelt und begründet werden kann.

c) Präzisierung der Problemformulierung: Konzeptspezifikation und Operationalisierung

Am Ende der Vorstudienphase sollen die zu beantwortende Forschungsfrage präzise formuliert werden. Es sind jene Konzepte (Begriffe, Theorieteile) fest-

⁷ Für eine intensive Literatursuche stehen inzwischen auch mehrere Datenbanksysteme zur Verfügung, z.B. bei dem Informationszentrum Sozialwissenschaften in Bonn. Oftmals verfügen auch die Universitätsbibliotheken über EDV erfaßte Bestandskataloge. Diese können nach Autorennamen, Schlag- und/oder Stichwörtern abgefragt werden.

zulegen, mit deren Hilfe die Fragestellung beantwortet werden soll. In diese Phase fallen auch erste Überlegungen zur Auswahl der geeigneten Forschungsmethoden. So entstehen Leitideen für den später auszuarbeitenden Forschungsplan. Ein weiterer Schritt besteht in der Überprüfung der konkreten Forschungsmöglichkeiten; dies ist die Aufgabe der *Explorationsphase*. Im Rahmen einer *Feldbegehung* sucht der Forscher Orte, Institutionen, Archive oder Situationen auf, in denen später die Datenerhebungen vorgenommen werden sollen.

Auf der Grundlage des in der Explorationsphase erworbenen Wissens wird der endgültige Forschungsplan ausgearbeitet. Da in dieser Phase von den forschungsleitenden Konzepten (Begriffen, Theorien, Theorieausschnitten) ausgegangen wird, heißt diese Arbeitsphase auch *Konzeptualisierungsphase*. In ihr sind folgende Teilaufgaben zu bewältigen:

- Das theoretische Vorverständnis vom Forschungsgegenstand grenzt zunächst zusammen mit dem Ziel der Untersuchung die Bereiche des betrachteten Realitätsausschnittes ein.
- Theoretische Vorstellungen sind die Grundlage für die gedankliche Strukturierung des Objektbereichs (dimensionale Analyse; Entscheidung über die für den Untersuchungsansatz relevanten Sachverhalte),
- die abbildende Beschreibung des Objektbereichs mit Hilfe sprachlicher Zeichen (Begriffe; Definitionen) und
- die Möglichkeiten der Herstellung der Beobachtbarkeit dieser begrifflich bezeichneten Sachverhalte (Auswahl und Begründung von Indikatoren; Operationalisierung).

Auf der Grundlage der theoretischen Vorstellungen werden die Aspekte oder »Dimensionen« des empirischen Untersuchungsgegenstandes ausgefiltert, die für die aktuelle Fragestellung als besonders bedeutsam erscheinen. Welche »Dimensionen« des betrachteten Realitätsausschnitts sind durch die spezifische Problemstellung explizit angesprochen?

Ziel der *dimensionalen Analyse* ist die Aufstellung eines Begriffssystems der Dimensionen des betrachteten historischen Realitätsausschnitts. Es gilt die empirische Wirklichkeit mit Begriffen zu verknüpfen (Herstellung einer Korrespondenz zwischen empirischen Sachverhalten und sprachlichen Zeichen). Bei der dimensional Analyse des Untersuchungsgegenstandes geht es darum

- herauszufinden, welche Aspekte an dem Untersuchungsgegenstand prinzipiell festgestellt werden können und
- zu entscheiden, welche Aspekte für die konkrete Untersuchung so bedeutsam sind, daß dazu Daten (Indikatoren) erhoben werden sollen.

Für beide Aufgaben ist es hilfreich, wenn auf bereits bewährte Theorien (oder Theorieausschnitte) zurückgegriffen werden kann.

Die Korrespondenz zwischen empirischen Sachverhalten und sprachlichen Zeichen wird mit Hilfe von *Definitionen* hergestellt, in denen Aspekte der

Realität durch sprachliche Zeichen (= Begriffe) repräsentiert werden. Begriffe und Dimensionen beziehen sich aufeinander. Nach Zetterberg soll unter Dimension eine »Eigenschaft der Wirklichkeit« des Gegenstandsbereichs verstanden werden; Begriffe dagegen sind Bestandteil der Sprache, mit denen der Gegenstandsbereich bezeichnet wird (vgl. Zetterberg, H.L., 1973: Theorie, Forschung und Praxis in der Soziologie, in: König, R. (Hrsg.), 1973: Handbuch der empirischen Sozialforschung, Bd. 1, 3. A., Stuttgart, S. 103–160).

»Der Definition hat also entweder eine dimensionale Analyse voranzugehen; d.h. die für die Diagnose relevanten Aspekte der zu untersuchenden Situation werden herausgearbeitet und mit Begriffen (mit sprachlichen Symbolen) 'bezeichnet'. Oder der für die Untersuchung geltenden Definition hat eine semantische Analyse voranzugehen, d.h. es wird geklärt, in welcher Bedeutung die interessierenden Begriffe im Untersuchungszusammenhang verwendet werden und welche Aspekte der Realität den Begriffen entsprechen (sollen)« (Kromrey, H., 1991: Empirische Sozialforschung, 5. überarb. u. erw. A., Opladen, S. 70). Die semantische Analyse von Begriffen hat somit die Frage zu beantworten, welche Bedeutungen sprachlichen Zeichen zugeschrieben werden. Die Begriffe werden in ihre Bedeutungskomponenten zerlegt; es wird festgestellt, auf welche Bedeutungskomponenten sich der theoretische Begriff bezieht. Dieses Begriffssystem ist der Orientierungsrahmen für die sich anschließenden Forschungsschritte.

Präzisierung der Fragestellung, dimensionale bzw. semantische Analyse sind gewichtige Schritte im Forschungsprozeß, da Festlegungen auf dieser Stufe die späteren Schritte in hohem Maß beeinflussen.

Unter *Dimensionen* sind somit diejenigen Einzelheiten zu verstehen, die an einem empirischen Sachverhalt unterschieden werden können. Eine empirische Untersuchung setzt ferner eine Präzisierung der zur Erklärung verwendeten Konzepte und Begriffe voraus. Wenn geklärt ist, welche Aspekte eines theoretischen Begriffs berücksichtigt werden sollen, bleibt die Frage zu klären, wie den theoretischen Begriffen beobachtbare Sachverhalte (*Indikatoren*) zugeordnet werden können, so daß eine empirische Messung im Rahmen der Datenerhebung möglich ist. Theoretische Begriffe sind bewußt so allgemein und umfassend gehalten, daß sie mehr als eine eng abgrenzbare Menge konkreter Ereignisse umgreifen. Es ist zu präzisieren, welche spezifischen (empirischen) Sachverhalte mit den verwendeten theoretischen Begriffe gemeint sind (Festlegung des empirischen Bedeutungsgehaltes).

Als *Operationalisierung* bezeichnet man dann die Angabe, wie einem theoretischen Begriff beobachtbare (d.h. erhebungstechnisch »meßbare«) Indikatoren zugeordnet werden. Die konkrete Vorgehensweise bei der Operationalisierung ist abhängig vom empirischen Bezug des zu operationalisierenden Begriffs. Bei Begriffen mit direktem Bezug lassen sich die durch den Begriff bezeichneten Sachverhalte unmittelbar beobachten bzw. wahrnehmen. Bei Begriffen mit indirektem empirischen Bezug müssen zunächst Indikatoren gebil-

det werden. Indikatoren sollen durch empirisch feststellbare Sachverhalte auf das Vorhandensein der nicht unmittelbar beobachtbaren, mit dem Begriff bezeichneten Sachverhalte verweisen. Korrespondenzhypothesen setzen – von theoretischen Begriffen ausgehend – Begriffe mit konkreten Aspekten der Wirklichkeit in Beziehung, d.h. sie geben an, welche Aspekte der Realität in Form von meßbaren Indikatoren im Rahmen einer konkreten Untersuchung unter den Begriff subsumiert werden sollen. Es muß – mit anderen Worten – eine Korrespondenz zwischen empirischen Sachverhalten und theoretischen Begriffen hergestellt werden.

Die Gültigkeit der Indikatorenbildung hängt entscheidend davon ab, wie genau die durch den Indikator beobachtbaren Sachverhalte die mit dem Begriff bezeichneten Sachverhalte abbilden. Die Indikatorenbildung ist daher anhand einer sorgfältigen Indikatorenanalyse hinreichend zu begründen.

Operationalisierungen bestehen somit aus Angaben, wie Messungen für einen bestimmten Begriff vorgenommen werden können. Erst anhand der erhobenen Messungen lassen sich Aussagen über den Gegenstandsbereich und damit auch über die – vorläufige – Akzeptanz oder Verwerfung der zu prüfenden Hypothesen machen.

Zentrale Entscheidungen in dieser Phase des Forschungsprozesses betreffen schließlich die Wahl und die Konstruktion der Forschungsinstrumente bzw. Datenerhebungsmethoden. Die Wahl der Erhebungsmethode bestimmt sich aus dem Forschungsgegenstand, aus der Fragestellung und aus der Operationalisierung. In den Sozialwissenschaften existieren eine Reihe verschiedener Datenerhebungsmethoden, z.B. Interviews, Beobachtungen, Inhaltsanalysen, sog. nicht-reaktive Meßverfahren wie Akten-/Dokumentenanalysen usw. Die Konstruktion des gewählten Instruments geschieht nach den besonderen Erfordernissen des zu untersuchenden Bereichs.

Die Konstruktion von Erhebungsinstrumenten setzt die Festlegung der zu unterscheidenden Ausprägungen der Begriffe bzw. der korrespondierenden Indikatoren voraus. Indikatoren und ihre Ausprägungen bilden die sog. Merkmale (oder Variablen) einer empirischen Untersuchung. Kennzeichnend für diese Untersuchungsphase sind die Fragen der Gültigkeit und Zuverlässigkeit:

- Sind die Indikatoren und die gewählten Ausprägungen gültig, d.h. erfassen sie genau diejenigen Tatbestände und diejenigen Differenzierungen der Realität, die mit der Forschungs-Problemstellung gemeint sind?
- Sind die verwendeten Erhebungsinstrumente zuverlässig, d.h. führt ihre wiederholte Anwendung – auch von verschiedenen Forschern – unter gleichen Bedingungen zu gleichen Ergebnissen?

In der Konzeptualisierungsphase ist schließlich noch die Art des Forschungsprojektes (das Forschungsdesign) festzulegen: Art der Aufgabenstellung, Zahl der einzubeziehenden Untersuchungsebenen, Einzelfallstudie oder vergleichende Studie, synchrone oder diachrone Studie usw.

Vielfach werden in dieser Phase bereits die wichtigsten Wege der Datenanalyse festgelegt, vor allem die zu verwendenden statistischen Modelle. Gerade die Datenanalyse erlaubt schließlich die Erarbeitung der gesuchten Aussagen bzw. die Prüfung der forschungsleitenden Hypothesen und nur durch ein adäquates Erhebungsinstrument können die auszuwertenden Daten überhaupt in der für bestimmte Datenanalyseverfahren geeigneten Weise erhoben werden. Die Planung der Datenanalyse und die Entwicklung des Erhebungsinstruments bilden in den meisten Fällen eine nahtlose Einheit.

Von der Sorgfalt und Qualität der in der Konzeptualisierungsphase geleisteten Arbeit hängen Ertrag und Nutzen einer empirischen Studie vollständig ab. Fehler und Versäumnisse sind nicht mehr oder allenfalls mit hohem Zeit- und Arbeitsaufwand zu korrigieren.

d) Auswahl der Untersuchungsobjekte, bei denen die Merkmalsausprägungen gemessen werden sollen

Zu den notwendigen Entscheidungen in dieser Phase des Forschungsprozesses gehört zum einen die Festlegung der Merkmalsträger (Elemente des Gegenstandsbereichs, z.B. Personen, Gebiete, Zeitpunkte etc.), zum anderen die Beantwortung der Frage, ob sämtliche Untersuchungselemente (= *Grundgesamtheit*) oder nur einige ausgewählte Elemente (= *Stichprobe*) einbezogen werden sollen (oder einbezogen werden können). In der Forschungspraxis beinhaltet diese Phase vor allem eine exakte Definition des Gegenstandsbereiches, d.h. die Festlegung der problemspezifischen Grundgesamtheit. Soll lediglich eine Teilauswahl untersucht werden, stellt sich die Frage, ob besonders typische Elemente (Fälle) aus der Grundgesamtheit herausgegriffen werden sollen, oder ob eine *repräsentative* Auswahl erforderlich ist (d.h. eine Stichprobe, die als strukturtreue Abbildung der Grundgesamtheit angesehen werden kann).

Will man eine Teilmenge von Untersuchungselementen bestimmen, so benötigt man Verfahren, die es erlauben, eine repräsentative Auswahl von Elementen zu ermitteln (= *Auswahlverfahren*). Die Wahl des Auswahlverfahrens verbindet forschungs- und gegenstandsadäquate Notwendigkeiten mit den praktischen Durchführungsmöglichkeiten. Es entscheiden sich hier Fragen der Verallgemeinerbarkeit und der Repräsentativität sowie der Auswirkung des gewählten Stichprobenumfangs auf die Auswertungsmöglichkeiten der erhobenen Daten (Probleme der Datenanalyse). Erheblichen Einfluß üben praktische Durchführungsprobleme aus: Kosten, institutionelle Bedingungen und Zugangsmöglichkeit des Forschungsfeldes.

e) Datenerhebungsphase

Nachdem das Erhebungsinstrument entwickelt ist, können gemäß dem Auswahlplan die benötigten Daten erhoben werden. Bei umfangreichen empiri-

schen Studien wird der Forscher Mitarbeiter beschäftigen, welche mit der Durchführung der konkreten Feldarbeit betraut sind. Diese Mitarbeiter sind vor Beginn der Feldarbeit zu schulen und während der Feldarbeit zu betreuen und zu kontrollieren. Ferner sind folgende Aufgaben zu erfüllen:

- Die korrekte Durchführung des Auswahlplans ist zu überwachen. Ausfälle sind zu protokollieren; ferner ist festzustellen, mit welchem Effekt die tatsächlich untersuchte Auswahl aufgrund von Ausfällen verzerrt ist.
- Die Zuverlässigkeit der Datenerhebung ist zu kontrollieren.
- Das pro Untersuchungseinheit vorliegende Datenmaterial (*Datensatz*) ist einer Vollständigkeits-, Konsistenz- und Glaubwürdigkeitskontrolle zu unterziehen.

Die Daten sind nach der Feldarbeit für eine maschinelle Weiterverarbeitung (mit Hilfe der EDV) auf den Erhebungsunterlagen (*Datenblätter*) zu codieren. Die Codierung der Daten hat nicht zuletzt ökonomische Gründe, da sie Arbeitszeit gegenüber Einträgen im Klartext spart. Im Falle numerischer Daten ist daher zunächst ein Codeplan zu erstellen, der vorschreibt, welcher Datenwert für eine bestimmte Merkmalsausprägung steht.

f) Datenerfassung und -aufbereitung

Die erhobenen Daten müssen festgehalten und in einer spezifischen Art aufbereitet werden, d.h. sie sind in eine für die Analyse geeignete Form zu übersetzen. Eine Datensammlung muß auf eine bestimmte Weise strukturiert werden, bevor eine Auswertung möglich ist. Entscheidend ist die Frage, ob im Hinblick auf die Auswertung eine Übertragung auf einen Datenträger (für eine EDV-gestützte Datenverarbeitung) erforderlich ist. Soll die Auswertung quantitativ mit Hilfe der EDV erfolgen, ist zunächst eine externe Rohdatendatei zu erstellen. Die Daten werden von den Datenblättern auf einen maschinenlesbaren Datenträger übertragen (z.B. auf eine Diskette oder Festplatte, wenn ein PC eingesetzt wird). Bei kleineren Datenmengen genügt es, wenn man einen Editor oder ein Textverarbeitungsprogramm nutzt, um die Daten aus den Urbelegen zeilenweise – entsprechend dem Codeplan – einzugeben. Bei Verwendung eines Textverarbeitungsprogramms ist darauf zu achten, daß keine Steuerzeichen zur Formatierung in der Rohdatendatei gespeichert werden, d.h. die Rohdatei ist im sog. ASCII-Format zu speichern. Diese Rohdatendatei wird im nächsten Arbeitsschritt in ein Statistik-Programmpaket eingelesen. An die Erfassung und Eingabe schließt sich meist eine umfangreiche Datenbereinigung an, bei der Fehler in den Daten (z.B. durch Schreib-, Codier- oder Übertragungsfehler) gesucht und beseitigt werden.

g) Datenanalysephase: Anwendung statistischer Modelle und Verfahren

Die Datenanalyse in der historischen Sozialforschung besteht überwiegend aus dem Einsatz statistischer Methoden unter Verwendung der EDV – heute insbesondere durch den Einsatz von Mikrocomputern (PCs) – und speziellen Statistiksoftware-Programmpaketen (den Datenanalysesystemen). Die Aufgabenbereiche der Statistik lassen sich in beschreibende (deskriptive) Statistik und schließende (Inferenz-) Statistik einteilen. Inhaltlich bestehen die Aufgaben der Statistik allgemein darin, die in den erhobenen Daten enthaltenen Informationen neu zu organisieren, um auf dieser Grundlage statistische Aussagen zu ermöglichen. Dazu stellt die Statistik eine Vielzahl von Modellen zur Verfügung. Formal lassen sie sich danach unterscheiden, wieviele Merkmale simultan in ein Modell einbezogen werden. Bezüglich der Datenanalyse lassen sich zunächst drei Unterscheidungen treffen:

– *Deskriptive, univariate Datenanalyse:*

Es geht um die Beschreibung bestimmter Sachverhalte bzw. die Überprüfung von Beschreibungen. Ein Untersuchungsgegenstand wird hinsichtlich einer Dimension (bzw. deren empirische Repräsentation, einer Variablen oder eines Merkmals) mit ihren Ausprägungen allein beschrieben. Auch mehrere Dimensionen können nebeneinander betrachtet werden, aber ohne einen Zusammenhang bzw. eine Abhängigkeit zwischen ihnen herzustellen. Zur Darstellung dieser Information wird diese mittels verschiedener Darstellungstechniken wie Graphiken, Tabellen oder Methoden (z.B. statistische Maßzahlen) verdichtet.

– *Deskriptive, bivariate Datenanalyse:*

In dieser Analysephase werden zweidimensionale Kreuztabellen erstellt, um einen Zusammenhang (oder eine Abhängigkeit) zwischen jeweils zwei Merkmalen untersuchen zu können. Bei diesem Arbeitsschritt werden auf der Grundlage von zweidimensionalen Häufigkeitsverteilungen Zusammenhangsmaße berechnet (d.h. statistische Maßzahlen, die Auskunft geben über den Grad eines Zusammenhangs oder einer Abhängigkeit zwischen jeweils zwei Merkmalen).

– *Deskriptive, multivariate Datenanalyse:*

In dieser Analysephase werden simultan mehr als zwei Merkmale in einem geeignet zu wählenden Analysemodell betrachtet. Das Gebiet der multivariaten Datenanalyse ist am weitesten ausdifferenziert; zwischen einer Vielzahl von Alternativen muß der Anwender zunächst sorgfältig entscheiden, welches Modell für seine Fragestellung geeignet ist und inwieweit seine Daten die Modellannahmen erfüllen. Nur wenn das gewählte Modell bezüglich beider Voraussetzungen adäquat ist, lassen sich die Ergebnisse sinnvoll interpretieren.

Die Aussagen sozialwissenschaftlicher Theorie beziehen sich auf eine Vielzahl empirischer Objekte, die in der Regel nicht in ihrer Gesamtheit erfaßt werden können. Aus der Grundgesamtheit (Menge aller Objekte), an denen die interessierenden Merkmale festgestellt werden können, wird daher eine Stichprobe

gezogen, die ein unverzerrtes, repräsentatives Bild der Grundgesamtheit garantiert. Unter dieser Voraussetzung lassen sich inferenzstatistische Verfahren anwenden, die zwei Zielsetzungen verfolgen können:

– *Schätzung von Vertrauens- oder Konfidenzintervallen:*

Für die Maßzahlen der deskriptiven Statistik können unter den Bedingungen der Zufallsauswahl mittels inferenzstatistischer Verfahren Grenzen angegeben werden, innerhalb derer der korrespondierende Grundgesamtheitsparameter mit einer bestimmten Wahrscheinlichkeit zu finden ist.

– *Verifizierende, Hypothesen testende Analyse:*

Theorien bestehen u.a. aus Aussagen über Zusammenhänge bzw. Abhängigkeiten zwischen ausgewählten Merkmalen des Gegenstandsbereichs der Theorie. Hypothesentestende Datenanalyse versucht, Erklärungen für bestimmte Sachverhalte mit Hilfe von Signifikanztests zu prüfen. Die in Hypothesen formulierten Aussagen beziehen sich auf die Grundgesamtheit, zu ihrer Prüfung stehen aber lediglich Stichproben zur Verfügung. Unter Rückgriff auf theoretische Verteilungsmodelle läßt sich aber prüfen, ob die in einer Hypothese formulierte Aussage für die Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit Gültigkeit beanspruchen kann oder nicht. In der Statistik sind bestimmte Regeln aufgestellt worden, wie die Entscheidung über Akzeptanz oder Ablehnung einer Hypothese zu fällen ist. Formal wird das Problem so gefaßt, daß zusätzlich eine zweite, konkurrierende Hypothese eingeführt wird, die sog. Nullhypothese. Während in der Ausgangshypothese (im Kontext des Signifikanztests auch Alternativhypothese genannt) z.B. ein bestimmter Zusammenhang zwischen zwei Merkmalen in der Grundgesamtheit angenommen wird, behauptet die Nullhypothese, daß der empirisch ermittelte Zusammenhang in der Stichprobe nur zufällig ist (für die Grundgesamtheit daher kein Zusammenhang angenommen werden kann). Die Nullhypothese wird nun mit Hilfe des Signifikanztests überprüft. Ergibt der Test, daß das in der Nullhypothese behauptete Ergebnis (z.B. »kein Zusammenhang zwischen den Merkmalen A und B«) sehr unwahrscheinlich ist, wird diese verworfen und stattdessen die Alternativhypothese akzeptiert, d.h. ihre Gültigkeit angenommen. Ein derartiges Testergebnis besagt also nicht, daß die Alternativhypothese richtig ist, sondern nur, daß die Nullhypothese mit hoher Wahrscheinlichkeit falsch ist.

Mit Hilfe von statistischen Verfahren kann geprüft werden, ob die theoretisch vorhergesagten Beziehungen zwischen ausgewählten Merkmalen in den erhobenen Daten nachweisbar sind oder nicht. Dieser Auswertungsschritt stellt die eigentliche Rückkopplung zwischen Theorie (und daraus abgeleiteten Hypothesen) und empirischen Resultaten her: Die empirischen Ergebnisse (festgeschrieben in »Beobachtungsaussagen«) sind – unter Berücksichtigung all der Einschränkungen, die bei der Operationalisierung, der Datenerhebung und -auswertung möglicherweise in Kauf zu nehmen waren – wieder zurückzuüberset-

zen. Die aufgestellten Hypothesen können dabei entweder bestätigt oder verworfen, aber auch ergänzt, spezifiziert, präzisiert oder revidiert werden. Man kann hier verschieden enge oder weite Ergebnisinterpretationen unterscheiden. Eine enge Interpretation besteht in der Prüfung, ob die empirischen Befunde geeignet sind, eine Hypothese (oder ein Hypothesenbündel) zu bestätigen oder nicht. Diese Interpretation hält sich eng an die Daten und sagt nichts darüber hinaus aus. Etwas weiter geht die Interpretation, wenn sie auch die Aussagefähigkeit der Daten, den Datengenerierungsprozeß mit einbezieht. Dies ist zum richtigen Verständnis von empirischen Befunden und zur Einordnung in vergleichbare Untersuchungen sehr bedeutsam, denn die Interpretation der aus den erhobenen Daten gewonnenen Forschungsergebnisse schließt die Entscheidung mit ein, welches Gewicht der Forscher den Ergebnissen der Datenanalyse zumißt und wie weit sie inhaltlich ausreichen, die Forschungshypothesen zu belegen bzw. wie weit die Theorie des Forschungsgegenstandes modifiziert werden muß.

Teil 2: Quantitative Methoden in der historischen Sozialforschung

1. Messen und Variablenbildung

1.1 Das »Strukturmodell« empirischer Forschung

Die Methodologie empirischer Forschung ist bisher am Modell des Forschungsprozesses mit voneinander abgrenzbaren und zeitlich aufeinander folgenden Arbeitsschritten dargestellt worden. Diese Perspektive ist in den Teilaspekten der folgenden Abschnitte – Merkmale und Messen, Auswahlverfahren, Datenanalyse – nicht mehr aufrechtzuerhalten. Die Entscheidungen über geeignete Erhebungsinstrumente, über den anzustrebenden Differenzierungsgrad der Daten, über das geeignete Auswahlverfahren von Untersuchungsobjekten und die Verwendung von Datenanalyseverfahren können weder sukzessive noch unabhängig voneinander getroffen werden. In der Literatur ist ein »Strukturmodell für nicht-experimentelle Forschungsprojekte« vorgeschlagen worden (vgl. Kromrey, H., 1991: Empirische Sozialforschung, 5. überarb. u. erw. A., Opladen, S. 141–143), das zum einen die Zusammenhänge zwischen den einzelnen Arbeitsschritten aufzeigt, zum anderen einen wichtigen Konkretisierungsschritt auf dem Wege zur kontrollierten Erhebung empirischer Daten verdeutlicht (s. Abbildung 1.1).

»Diese Konkretisierung bezieht sich zum einen auf den *Grad der Differenzierung*, mit dem die als Indikatoren benutzten empirischen Sachverhalte gemessen werden sollen. Der Grad der Differenzierung kann von der bloßen Feststellung des Vorhandenseins oder Nichtvorhandenseins eines Tatbestandes bis zur möglichst präzisen (in metrischen Maßeinheiten angebbaren) Erfassung des Ausmaßes seines Vorhandenseins reichen. Durch die Angabe des Differenzierungs- oder Variabilitätsgrades, der bei der Datenerhebung berücksichtigt werden soll, werden die begrifflich bezeichneten Indikatoren zu *Variablen*.

Die Konkretisierung bezieht sich zum anderen auf die Regeln zur Feststellung der zu messenden Sachverhalte an den zu untersuchenden Objekten, d.h. auf die *Meßvorschriften*. Die interessierenden Sachverhalte sollen nicht weiter als abstrakte gedankliche Konzepte betrachtet werden, sondern als Merkmale von Untersuchungsobjekten. Zum Beispiel kann der mit dem Begriff 'Einwohnerzahl' bezeichnete Sachverhalt nicht als solcher, sondern nur als Merkmal konkreter Dörfer oder Städte oder Länder oder Staaten gemessen werden. Ebenso ist 'sozialer Status' nicht als abstraktes, begrifflich definierbares Phänomen empirisch feststellbar, wohl aber als Merkmal von Personen. Meßvorschriften sind also Regeln, nach denen abstrakte gedankliche Konzepte (wie

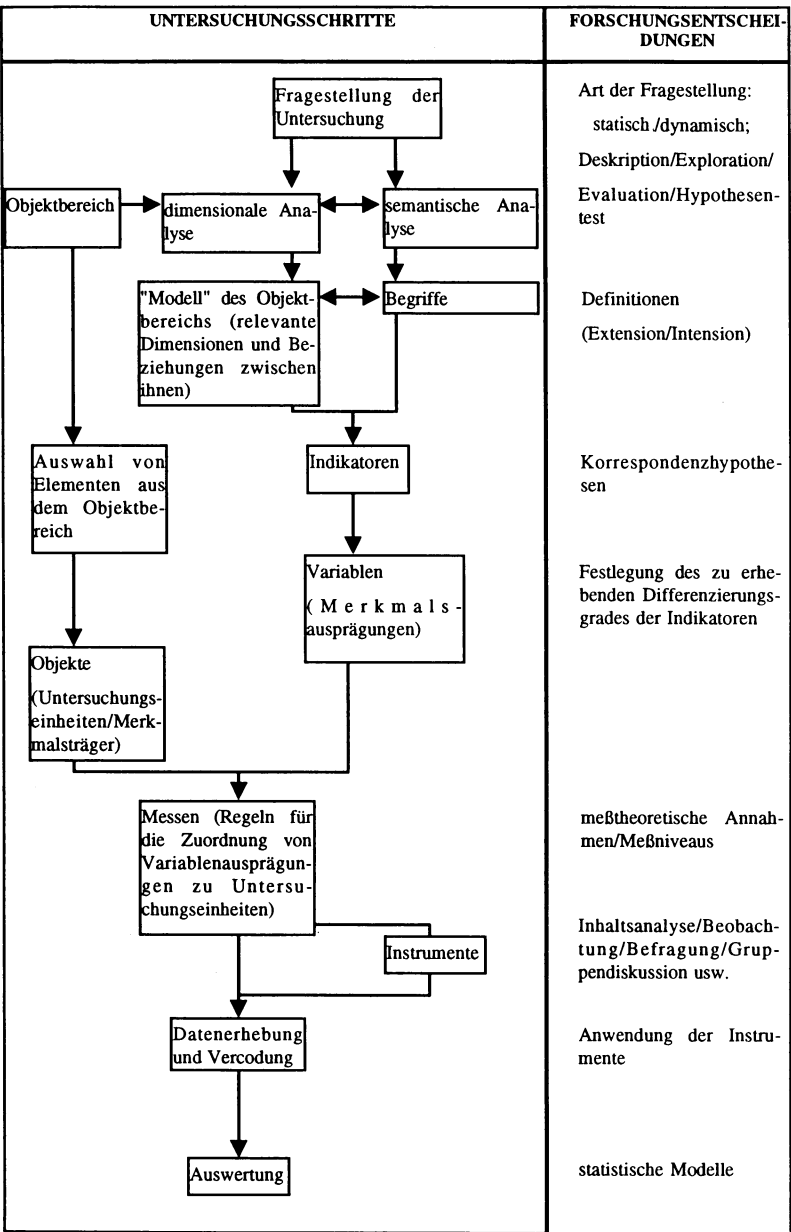


Abb. 1.1: "Strukturmodell" für nicht-experimentelle Forschungsprojekte

Quelle: Kromrey, H., 1991: Empirische Sozialforschung, 5. überarb. u. erw. A., Opladen, S. 142.

Körpergröße, Gewicht, Geschwindigkeit, Farbe, Alter, Einkommen, Bildung ...) an konkreten empirischen Objekten festgestellt werden sollen. Mit anderen Worten: Sie geben an, auf welche Weise der Forscher seine gedanklichen Konzepte mit der Realität zu verknüpfen beabsichtigt« (Kromrey, a.a.O., S. 143).

Aus der Erhebungsphase gehen die Rohdaten hervor. Sie werden aufbereitet und mit Hilfe statistischer Modelle analysiert. Die Konkretisierung bezieht sich hier auf die zur Auswahl stehenden Modelltypen. Der Ausdruck *Modell* bezeichnet allgemein die symbolische Darstellung der Struktur von Sachverhalten unter bestimmten Gesichtspunkten. Das Modell bildet einen Ausschnitt aus der Wirklichkeit nach, wählt Bestandteile aus, die für den Realitätsausschnitt repräsentativ sind. Die realen Sachverhalte werden abgebildet durch Merkmale; statistische Modelle bilden die Struktur realer Sachverhalte ab, und zwar durch statistische Kennzahlen (z.B. Mittelwerte, Streuungen etc.) und in Form von Zusammenhängen zwischen den erhobenen Merkmalen (z.B. in Form von Maßzahlen bivariater Assoziation). Verschiedene Modelle, angewandt auf dieselbe Ausgangsstruktur, verarbeiten und betonen unterschiedliche Informationen, beantworten verschiedene Fragen. In der Statistik gibt es verschiedene Modelltypen. Es gibt Modelle für Schlüsse von Stichproben auf Grundgesamtheiten; es gibt Modelle, die Beziehungen zwischen Merkmalen repräsentieren, und Modelle, die eine Reduktion der vielfältigen Informationen von Daten auf einige wenige, besonders relevante Größen beinhalten.

Konkretisierung durch Anwendung statistischer Modelle zur Informationsgewinnung und Verdichtung ist an bestimmte Voraussetzungen gebunden:

- »Statistische Modelle implizieren im Vergleich zur abgebildeten Realität erhebliche Vereinfachungen und Formalisierungen. Wenn man das Ziel hat, die unübersichtliche Vielfalt der in einer Datenbasis enthaltenen Informationen zu verdichten, Komplexität zu reduzieren, dann geht das nur, indem man wesentliche von unwesentlichen Informationen trennt und die wesentlichen Informationen hervorhebt. Was allerdings 'wesentlich' oder 'unwesentlich' sein soll, läßt sich nur von der Fragestellung der Untersuchung her entscheiden.
- Statistische Modelle sind nur auf zähl- und meßbare Tatbestände anwendbar. Da aber 'Meßbarkeit' nicht mit 'Quantifizierbarkeit' identisch ist ... wird es im wesentlichen eine Frage der Operationalisierung sein, ob und inwieweit damit Einschränkungen ihrer Anwendbarkeit auf soziale Sachverhalte verbunden sind.
- Statistische Modelle beziehen sich niemals auf den Einzelfall, sondern immer auf Gruppen vergleichbarer Fälle, auf eine irgendwie abgegrenzte *Objektmenge*« (Kromrey, a.a.O., S. 150).

Die Statistik selbst ist lediglich »Werkzeug« (vgl. Harder, Th., 1974: Werkzeug der Sozialforschung, München) des Sozialforschers; ihre Ergebnisse können niemals besser sein als die Daten, auf die statistische Modelle angewendet werden.

1.2 Grundbegriffe der Statistik

Vor jeder Datenerhebung muß zunächst eindeutig festgelegt werden, über welche Gegenstände oder Personen bestimmte Aussagen gemacht werden sollen (z.B. erwerbstätige Personen). Je nach Forschungsgebiet und Zielsetzung wird eine Menge von Objekten, deren Massenerscheinungen mit den statistischen Methoden untersucht werden sollen, definiert bzw. abgegrenzt. Die zu untersuchende Menge von gleichartigen Elementen wird als *statistische Masse* bezeichnet (z.B. erwerbstätige Personen in der Bundesrepublik im Dezember 1994). In einer statistischen Masse sollten sinnvollerweise nur solche Elemente zusammengefaßt werden, die vom Untersuchungsziel her als gleichartig angesehen werden. Eigenschaften oder Merkmale, die angeben, ob ein beliebiges Objekt ein Element einer zu untersuchenden statistischen Masse ist oder nicht, werden *Identifikationsmerkmale* genannt (Beispiel: erwerbstätige Person, BRD, Dezember 1994 als Ausprägungen der Merkmale OBJEKT, LAND und ZEIT). Jede statistische Einheit ist Träger derselben Ausprägung eines sachlichen, räumlichen und zeitlichen Identifikationsmerkmals. Bei einer Abgrenzung der statistischen Massen in zeitlicher Hinsicht können zwei verschiedene Fälle unterschieden werden: Wird eine Masse zu einem bestimmten Zeitpunkt (Stichtag) erfaßt, so bezeichnet man sie als *Bestandsmasse*. Im Gegensatz zu den Bestandsmassen bezeichnet man als *Bewegungsmassen* diejenigen statistischen Massen, deren Einheiten innerhalb eines bestimmten Zeitraumes erfaßt werden können (z.B. Bruttosozialprodukt pro Jahr, Geburten, Todesfälle).

Die Gesamtmenge der Untersuchungseinheiten, die im Hinblick auf ein Untersuchungsziel dieselben Ausprägungen eines sachlichen, räumlichen und zeitlichen Identifikationsmerkmals tragen, wird Grundgesamtheit oder Population genannt. Statistische Einheiten können sein: Personen, Dinge, soziale Gebilde, Verwaltungs-/Wahlbezirke, Städte, Staaten, Zeiteinheiten (z.B. Monat, Jahr), Vorkommnisse oder Ereignisse, Handlungen o.ä.

Die Statistik dient im wesentlichen der quantitativen Erfassung und überschaubaren Aufbereitung von massenhaft auftretenden Einzelercheinungen. Diejenigen Eigenschaften der Elemente der statistischen Masse, die unmittelbar die forschungsrelevanten Tatbestände beschreiben und die somit unmittelbar Gegenstand der Erhebung sind, nennt man *Erhebungsmerkmale*, kurz: *Merkmale*. Die Erscheinungsform eines Merkmals bezeichnet man als *Merkmalsausprägung*. Ein Erhebungsmerkmal unterteilt eine Ausgangsmasse in verschiedene Teilmassen. So kann z.B. die Bevölkerung eines Landes als statistische Ausgangsmasse nach dem Merkmal Geschlecht gegliedert werden, wobei dieses Merkmal in zwei Ausprägungen auftritt: männliche und weibliche Bevölkerung. Eine Merkmals- oder Eigenschaftsdimension, die mit einem Begriff bezeichnet wird und mehrere Ausprägungen annehmen kann, wird als *Variable* bezeichnet (Variable = Begriff + (mindestens 2) Ausprägungen). Variablen können im Prinzip eine Vielzahl von Ausprägungen annehmen. Die Anzahl der möglichen Ausprägungen hängt dabei nicht in erster Linie von der statistischen

Einheit ab, die mit der Variablen beschrieben werden soll, sondern vor allem von der Differenziertheit der begrifflichen Strukturierung, von der Methode der Operationalisierung und des zur Verfügung stehenden Meßinstrumentariums. Welche Ausprägungen unterschieden werden sollen, ist eine weitere wichtige Entscheidung im Forschungsprozeß: Wie differenziert sollen Unterschiede im zu untersuchenden Gegenstandsbereich durch Messungen abgebildet werden?

Merkmale können nach unterschiedlichen Gesichtspunkten systematisiert werden. So kann z.B. unterschieden werden zwischen räumlichen, zeitlichen und sachlichen Merkmalen:

- Räumliche Merkmale bezeichnen den Beobachtungsort oder den Beobachtungsraum;
- zeitliche Merkmale kennzeichnen den Erhebungszeitpunkt oder -zeitraum;
- sachliche Merkmale beschreiben statistische Einheiten unabhängig von deren Vorkommen in Zeit und Raum.

Eigenschaften von statistischen Einheiten lassen sich ferner in *qualitative* und *quantitative* Merkmale klassifizieren. Die Merkmalsausprägungen quantitativer Merkmale sind reelle Zahlen. Kennzeichnend für quantitative Merkmale ist, daß sie in einer Maßeinheit verfügbar sind (z.B. Einkommen in DM, Körpergröße in cm, Gewicht in kg) und daß die Abstände zwischen den Ausprägungen angebar sind. Man spricht in diesem Fall auch von *metrischen Merkmalen*. Als qualitative Merkmale bezeichnet man Eigenschaften, die sich nicht in Zahlengrößen ausdrücken lassen; die Ausprägungen sind Wörter, meist feststehende Begriffe (z.B. Geschlecht, Beruf, Religionsbekenntnis usw.). Solche Begriffe, die statistische Einheiten (lediglich) in sich ausschließende Teilmengen zerlegen oder in Klassen gleicher Objekte einteilen, nennt man *klassifikatorische Merkmale*. Erlauben Merkmale nicht nur eine Klassifikation der statistischen Einheiten in sich ausschließende Teilmengen, sondern implizieren sie zusätzlich noch eine Rangordnung der Teilmengen (also einen Vergleich der Elemente hinsichtlich der Stärke oder Intensität eines Merkmals), dann handelt es sich um *komparative* Merkmale (Beispiel: Schulbildung).

Die Frage, welche Ausprägungen bei einer Variablen unterschieden werden sollen, und was diese Ausprägungen bezüglich der einzelnen Untersuchungseinheiten besagen, ist von zentralem Interesse. Die Festlegung des Wertebereiches einer Variablen und die Präzisierung des Aussagegehalts der Ausprägungen sind Gegenstand eines eigenständigen Bereichs der Statistik: der Meßtheorie und Skalierung. Ganz allgemein wird die Zuordnung von Zahlen zu Objekten, so daß die Zahlen bestimmte Eigenschaften der Objekte ausdrücken, Messung genannt (Messen als strukturtreue Abbildung). Genauer formuliert: Zuordnung von Ziffern zu Objekten entsprechend den Ausprägungen der betrachteten Merkmale. Das Meßergebnis ist demnach die symbolische Abbildung der empirischen Merkmalsausprägung. Diese Zuordnung muß systematisch, d.h. für alle Objekte gleich und nach gleichbleibenden Zuordnungsregeln durchgeführt werden. Sie gehorcht folgenden Gestaltungskriterien:

1. *Untersuchungsziel*: Die Merkmalsausprägungen müssen den Zweck der Untersuchung widerspiegeln.
2. *Einheitlichkeit*: Die Zuordnungen sollten nach einem einheitlichen Klassifikationsprinzip entwickelt sein. Das bedeutet, daß sie sich wirklich auf die eine Variable allein beziehen, deren Ausprägungen sie darstellen (Forderung nach Eindimensionalität).
3. *Eindeutigkeit*: Die Ausprägungen sollten sich wechselseitig ausschließen, damit eine eindeutige Zuordnung gewährleistet ist.
4. *Vollständigkeit*: Die Zuordnungen sollten erschöpfend sein, so daß keine Objekte ohne eine Merkmalsausprägung sind. Ansonsten wäre der zu untersuchende Realitätsausschnitt nur unzureichend erfaßt.

Wir beschäftigen uns nur mit solchen Variablen, deren Werte (reelle) Zahlen sind. Betrachten wir z.B. die Variable »Stellung im Beruf« mit den 5 Ausprägungen »Beamter«, »Angestellter«, »Arbeiter«, »Selbstständiger«, »sonstige Berufsstellungen«, so können wir diesen Ausprägungen die Zahlen 1 für »Beamter«, 2 für »Angestellter«, 3 für »Arbeiter«, 4 für »Selbstständiger«, 5 für »sonstige Berufsstellungen« zuweisen. Die Variable »Stellung im Beruf« kann also die Werte 1, 2, 3, 4, 5 annehmen. Variablen werden immer mit großen Buchstaben symbolisiert, z.B. X, Y, Z, ihre Werte mit kleinen Buchstaben, z.B. x_i , y_i , z_i , wobei

x_i = Wert der Variablen X für die Untersuchungseinheit i ist ($i = 1, \dots, n$),

y_i = Wert der Variablen Y für die Untersuchungseinheit i ist ($i = 1, \dots, n$),

z_i = Wert der Variablen Z für die Untersuchungseinheit i ist ($i = 1, \dots, n$),

mit n gleich dem Stichprobenumfang.

Die Tatsache, daß Ziffern nach unterschiedlichen Regeln zugeordnet werden können, führt zu verschiedenen Skalenarten:

- Von einer *Nominalskala* (von nominalem Meßniveau) spricht man, wenn von den Beziehungen (Relationen) zwischen den Ziffern der Meßskala nur die Gleichheit/Ungleichheit empirisch interpretiert werden darf. Die Ausprägungen entsprechen dann verschiedenen Zuständen, Situationen usw. Die zugeordneten Zahlen repräsentieren lediglich die Gleichheit bzw. Ungleichheit der Objekte hinsichtlich des gemessenen Merkmals. Die zugeordneten Zahlen stellen keine Größenangaben dar, sondern nur eine andere Art der Bezeichnung oder Namensgebung. Man könnte schließlich die Ausprägungen anstatt durch Zahlen durch andere Symbole – z.B. durch Buchstaben – kennzeichnen. Geschlecht ist ein Beispiel für eine Dimension, die nur auf Nominalskalenniveau gemessen werden kann. Man kann Personen nach dem Geschlecht in weibliche oder männliche Personen einteilen. Es existiert aber kein sachliches Kriterium, zwischen Männern und Frauen eine Ordnung herzustellen. Entsprechend ist es völlig willkürlich, ob man den Männern die Zahl 1 und den Frauen die Zahl 2 oder umgekehrt zuordnet.

Ebensogut könnte man 0 und 1 oder eine beliebig andere Zahlenkombination verwenden. Die Zahlen stellen lediglich kürzere Bezeichnungen dar, sie können aber nicht als Zahlen im üblichen Sinn interpretiert werden (z.B. »2« ist größer als »1«).

- Eine *Ordinalskala* (ordinales Meßniveau) liegt vor, wenn von den Beziehungen zwischen den Zahlen der Meßskala neben der Gleichheit/Ungleichheit auch die Rangordnung (größer/kleiner) empirisch interpretiert werden darf. Die Rangordnung gibt allerdings keinen Aufschluß darüber, wie groß der Unterschied zwischen zwei Rangplätzen ist. Das bedeutet vor allem, man kann keine sinnvollen Differenzen bei rangskalierten Daten bilden. Ein komparatives Merkmal ist z.B. die »formale Bildungsqualifikation« (= höchster Schulabschluß) mit den Ausprägungen: 1 »Hauptschulabschluß«, 2 »Mittlere Reife«, 3 »Abitur«, 4 »Fachhochschulabschluß«, 5 »Universitätsabschluß«, 0 »kein Abschluß (trotz beendeter Schulzeit)«.
- Eine *Intervallskala* ist gegeben, wenn von den Beziehungen zwischen den Zahlen der Meßskala auch die Abstände interpretierbar sind. Dadurch werden Aussagen möglich wie »Element A ist um x Einheiten verschieden von Element B« (z.B. größer, höher, kleiner, schlechter usw.). Intervallskalierte Daten besitzen keinen absoluten Nullpunkt, es wird höchstens ein Nullpunkt nach Übereinkunft festgelegt. Dadurch lassen sich keine interpretierbaren Multiplikationen und Divisionen von Variablenwerten durchführen; insbesondere lassen sich keine Aussagen machen wie »Der Variablenwert des Elementes A ist x-mal so groß wie derjenige des Elements B«. Die in ° C gemessene Temperatur ist intervallskaliert. Ein Temperaturwert von 30 ° C bedeutet im physikalischen Sinne nicht, daß es doppelt so warm ist wie bei 15 ° C.
- Von einer *Ratioskala* (Verhältnisskala) spricht man, wenn zusätzlich noch der Nullpunkt als Meßskala eine empirische Bedeutung hat und wenn dementsprechend auch die Größenverhältnisse zwischen den Zahlen (engl.: ratio) als Quotienten zwischen den Ausprägungen aussagefähig sind. Ein ratskaliertes Merkmal ist z.B. das Einkommen. Dem Skalenwert (d.h. dem Meßwert) »0 DM« entspricht in der Realität die Ausprägung »überhaupt kein Einkommen«.

Quantitative Merkmale können durch die Bildung von Intervallen in Kategorien eines ordinalen Merkmals überführt werden (z.B. das Merkmal KÖRPERGRÖSSE: »klein«: <1,55 m; »mittel«: 1,55–1,77 m; »groß«: > 1,77 m).

Die Maßskalen zur Messung der Variablen sind für die Anwendung statistischer Methoden von grundlegender Bedeutung, da viele Verfahren nur auf Daten eines bestimmten Skalen- oder Meßniveaus anwendbar sind. Die Skalenniveaus bilden in der genannten Reihenfolge eine hierarchische Stufung mit zunehmenden Informationsgehalt. Eine Variable mit einem bestimmten Skalenniveau kann auch in eine Variable mit einem niedrigeren Skalenniveau umgewandelt werden. Der umgekehrte Schritt ist dagegen nicht möglich. Das Pro-

blem der Meßniveaus beruht darauf, daß beobachtete Daten häufig ein Meßniveau haben, das nur in einem numerischen Relativ abbildbar ist, das nicht alle Eigenschaften reeller Zahlen besitzt. Das Meßniveau ist davon abhängig, welche der vier folgenden Eigenschaften reeller Zahlen tatsächlich erreicht werden:

- a) Kann man Objekte als gleich oder ungleich unterscheiden?
- b) Kann man Objekte in eine eindeutige Ordnung bringen?
- c) Ist es möglich, gleiche Abstände zwischen Kategorien anzunehmen?
- d) Existiert ein absoluter Nullpunkt?

Diese vier Kriterien bauen hierarchisch aufeinander auf, so daß ein höheres Kriterium immer die Existenz des niedrigeren Kriteriums voraussetzt. Damit liegt auch eine Hierarchie von Meßniveaus von niedrigem zu höherem vor. Die Tabelle 1.1 zeigt die Zusammenhänge im Überblick.

Tab. 1.1: Überblick über die Meßniveaus

Meßniveau	Mögliche empirische Aussagen	Beispiele
Nominal (Klassifikatorisch)	Unterscheidung: 1. Gleichheit und Ungleichheit	Konfessionszugehörigkeit, Geschlecht, Familienstand, Parteizugehörigkeit
Ordinal (Komparativ)	Unterscheidung und Rang: 1. Gleichheit und Ungleichheit 2. Rangfolge	Schulabschluß, Schulnoten, Altersklassen, soziale Schichtung
Intervall (Metrisch)	Unterscheidung, Rang, Abstand: 1. Gleichheit und Ungleichheit 2. Rangfolge 3. Gleichheit von Differenzen	Temperaturmessung in °C, Kalenderzeitrechnung,
Ratioskala (Verhältnisskala, mit absolutem Nullpunkt)	Unterscheidung, Rang, Abstand, Verhältnis: 1. Gleichheit und Ungleichheit 2. Rangordnung 3. Gleichheit von Differenzen 4. Gleichheit von Quotienten	Gewicht, Körpergröße, Einkommen Wahlergebnisse in Prozent

Merkmale lassen sich nicht nur nach ihren Beziehungen zwischen ihren numerischen Ausprägungen klassifizieren, sondern auch nach deren Anzahl. Grob unterscheidet man in der Statistik zwischen *diskreten* und *stetigen* Merkmalen:

- Ein diskretes Merkmal liegt vor, wenn die Anzahl der möglichen Merkmalsausprägungen abzählbar ist, d.h. es existiert eine Anordnung der Ausprägungen, entlang derer diese nacheinander aufgezählt werden können. Die Merkmalsausprägung beruht auf einem Zählvorgang (z.B. Kinderzahl: 0, 1, 2, ...; Haushaltsgröße: 1, 2, ...). Innerhalb der Gruppe der diskreten Merkmale lassen sich endliche und unendliche Merkmale unterscheiden. Im einfachsten Fall besitzt ein Merkmal lediglich zwei Ausprägungen (häufig bei nominalskalierten Merkmalen; z.B. Eigenschaft vorhanden oder nicht vorhanden); es heißt dann *dichotomes* Merkmal. Auf der anderen Seite sind bei den unendlichen Merkmalen die natürlichen Zahlen abzählbar entlang der Folge 1, 2, ..., ∞ .
- Ein stetiges Merkmal liegt vor, wenn die Anzahl der Merkmalsausprägungen nicht mehr abzählbar ist. Die Merkmalsausprägung beruht auf einem Meßvorgang. Stetig sind häufig intervall- oder verhältnisskalierte Merkmale (z.B. Gewicht: .. kg; Größe: .. m). Charakteristisch für diese Merkmale ist, daß ihre Ausprägungen jede mögliche reelle Zahl annehmen können.

»Messen« – im hier gemeinten Sinn als strukturtreue Abbildung – ist also die Zuordnung von Symbolen zu Sachverhalten, wobei den zu berücksichtigenden Unterschieden in den Sachverhalten auch Unterschiede in den Symbolen entsprechen müssen und wobei den Beziehungen zwischen den Sachverhalten auch Beziehungen zwischen den Symbolen entsprechen müssen (das Umgekehrte gilt nicht).

Im Vorgriff auf das Thema »statistische Auswertung« sei im Zusammenhang mit dem Skalenniveau von Variablen vorausgeschickt: die Statistik stellt für unterschiedliche Zwecke verschiedene mathematische Modelle der Informationsverarbeitung und -verdichtung zur Verfügung. Unter diesen ist das für die eigene Zielsetzung sachgerechteste Modell auszuwählen. Der Kreis der anwendbaren statistischen Methoden wird von dem Skalenniveau der Merkmale bestimmt. Es sind deshalb unter den verschiedenen Modellen jeweils die auszuwählen, die dem vorliegenden Datenniveau angemessen sind.

Für die Operation des Messens der Ausprägungen von Variablen sind zwei klassische Gütekriterien zu unterscheiden: *Gültigkeit* (»validity«) und *Zuverlässigkeit* (»reliability«). Gültig (oder valide) ist ein Merkmal dann, wenn es tatsächlich den Sachverhalt mißt (oder indiziert), der mit dem definierten Begriff bezeichnet worden ist, d.h. die im Begriff bezeichneten Sachverhalte abbildet. Als zuverlässig wird ganz allgemein eine Messung dann bezeichnet, wenn ihre wiederholte Anwendung unter gleichen Bedingungen gleiche Ergebnisse liefert. Bei genauerer Betrachtung kann man verschiedene Aspekte der Zuverlässigkeit unterscheiden (vgl. Kromrey, a.a.O., S. 184):

- Sie setzt zunächst *Objektivität* der Messung voraus, d.h. verschiedene Forscher müssen bei Vorliegen derselben Gegebenheit zu gleichen Ergebnissen kommen (auch intersubjektive Stabilität der Meßwerte genannt).

- Der zweite Aspekt beinhaltet die *intertemporale Stabilität* der Meßwerte: Bei wiederholter Messung desselben Sachverhalts erhält man die gleichen Ergebnisse. In der Regel wird dieser Aspekt als der Kernpunkt der Zuverlässigkeit angesehen (Zuverlässigkeit im engeren Sinne).
- Der dritte Aspekt, die *interinstrumentelle Stabilität* der Meßwerte, verweist darauf, daß die gleiche Merkmalsdimension durchaus mit Hilfe unterschiedlicher Instrumente gemessen werden kann.

Betrachten wir die Beziehung zwischen den beiden Gütekriterien Gültigkeit und Zuverlässigkeit: »Wenn man sagt, daß durch 'gültiges' Messen genau *das* in der Realität erfaßt wird, was mit den in der Forschung verwendeten Begriffen gemeint ist, dann kann die Gültigkeit niemals höher sein als die Zuverlässigkeit des verwendeten Meßinstruments. Denn in dem Maße, wie unterschiedliche Meßwerte *nicht* Unterschiede des gemessenen Merkmals zum Ausdruck bringen, sondern Resultat mangelnder Stabilität der Meßergebnisse sind, sind auch die Forschungsergebnisse insgesamt nicht 'gültig'; sie ist aber keine hinreichende Bedingung. Meßwerte können in noch so hohem Maße intertemporale, intersubjektive und interinstrumentelle Stabilität aufweisen, also 'zuverlässig' sein: sobald unangemessene ('falsche') Indikatoren ausgewählt und gemessen werden, sind die Resultate nicht gültig in bezug auf das, was mit den theoretischen Begriffen gemeint war. Das Meßinstrument mißt zwar zuverlässig, aber es mißt etwas anderes, als es messen soll« (Kromrey, a.a.O., S. 185).

2. Das Verhältnis von Grundgesamtheit und Stichprobe: Auswahlverfahren

2.1 Grundlagen der Stichprobenbildung

Die Statistik liefert Informationen über bestimmte Eigenschaften oder Merkmale, gemessen an Mengen von Untersuchungseinheiten. Ein weiterer Untersuchungsschritt besteht daher in der Angabe der Untersuchungseinheiten, an denen die interessierenden Merkmale empirisch festgestellt werden sollen (Festlegung des Objektbereichs: Welche Personen oder Sachverhalte sollen untersucht werden?). Die Festlegung des Objektbereichs erfolgt im allgemeinen in enger Verbindung mit der Festlegung der Gesamtmenge von Einheiten, über die durch eine statistische Untersuchung eine Aussage gemacht werden soll. Diese Gesamtmenge heißt Grundgesamtheit oder Population (»target population«). Gesamtheiten sind im allgemeinen endlich (vom Umfang N), wobei die Anzahl der Einheiten mitunter extrem hoch sein kann. Eine exakte Definition oder Abgrenzung der Grundgesamtheit ist zur Durchführung einer empirischen Untersuchung von großer Bedeutung: Aussagen einer Untersuchung gelten nur für die Objekte der Grundgesamtheit. Wie eine solche Abgrenzung vorgenommen wird, bestimmt sich hauptsächlich aus dem zugrunde liegenden Erkenntnisin-

teresse. Die Grundgesamtheit ist so zu wählen, daß sie für die Beantwortung der Forschungsfrage und die Überprüfung eventuell formulierter Hypothesen hinreichend geeignet ist. Anzustreben ist eine theoretisch gehaltvolle Populationsdefinition – unter der Voraussetzung, daß entsprechende präzisierte theoretische Annahmen vorliegen. Eine zu restriktive Definition der Grundgesamtheit kann dazu führen, daß über einen Großteil des Gegenstandsbereiches einer Theorie keine Aussagen mehr getroffen werden können.

Die Bestimmung der Grundgesamtheit hat neben der sachlichen Abgrenzung ferner eine 'extensionale' Dimension zu berücksichtigen: Jede zu untersuchende Gesamtheit muß zeitlich und räumlich eindeutig abgegrenzt werden. Ein Beispiel für eine Grundgesamtheit ist etwa »Sämtliche Haushalte in der BRD zum Zeitpunkt 24. August 1994«.

- *Zeitliche Abgrenzung:* Bei der Bestimmung des Zeitkriteriums sind die Einflußgrößen a) inhaltliche Relevanz, b) ökonomisches Vorgehen und c) Verfügbarkeit der Daten zu berücksichtigen. Ferner ist zu berücksichtigen, ob eine Grundgesamtheit zu einem bestimmten Zeitpunkt betrachtet werden soll, oder ob Aussagen über ein ganzes Zeitintervall gelten sollen.
- *Räumliche Abgrenzung:* Die Grundgesamtheit ist nach einem regionalen Abgrenzungskriterium zu definieren. Beispiele sind: Bevölkerung eines Bezirkes oder Landes, die Haushalte einer bestimmten Großstadt etc.

Wenn bisher aufgezeigt wurde, daß aufgrund eines spezifischen Erkenntnisinteresses eine Eingrenzung der zu untersuchenden Grundgesamtheit sinnvoll ist, so soll nicht der Eindruck erweckt werden, bei dieser Beschränkung handele es sich bereits um eine Auswahl im Sinne der Stichprobenbildung (auf die wir weiter unten ausführlich eingehen). Zwar ist jede Grundgesamtheit als Teilmenge einer umfassenderen Grundgesamtheit vorstellbar¹ und stellt insofern tatsächlich eine 'Auswahl' dar, jedoch ist die Zielsetzung im Vergleich zu den Auswahlen im Sinne der Stichprobenbildung streng zu unterscheiden. Bei der ersteren 'Auswahl' handelt es sich um eine inhaltlich begründete Eingrenzung des Objektbereiches, über die hinaus keine Aussagen getroffen werden sollen. Demgegenüber verfolgt die Stichprobenbildung tatsächlich eine solche Verallgemeinerung der Aussagen über die Zahl der in der Auswahl enthaltenen Elemente hinaus.

In einem zweiten Schritt ist festzulegen, ob die Datenerhebung für die Gesamtheit aller Untersuchungseinheiten erfolgen soll (= Voll- oder Totalerhebung) oder ob man sich auf eine Teilmenge der Grundgesamtheit (eine »Auswahl«) beschränken kann (= Teilerhebung). Die Vollerhebung ist dadurch charakterisiert, daß sämtliche Einheiten des zu erhebenden Sachverhalts mit den interessierenden Merkmalen erfaßt, gemessen und aufbereitet werden. Eine Vollerhebung ist in den meisten Fällen aus mehreren Gründen nicht möglich:

¹ In diesem Zusammenhang ist von »Supergesamtheiten« gesprochen worden (vgl. Cochran, W.G., 1972: Stichprobenverfahren, Berlin/New York).

- aus ökonomischen Gründen: Vollerhebungen sind in der Regel zu aufwendig, zu teuer und zu zeitraubend;
- aus Gründen der Zugänglichkeit: Eine Vollerhebung erweist sich technisch oder organisatorisch als nicht durchführbar;
- aus prinzipiellen Gründen: Möglicherweise sind gar nicht sämtliche interessierenden Untersuchungseinheiten erfaßbar.

Vollerhebungen sind vor allem dann möglich, wenn der Umfang der interessierenden Grundgesamtheit relativ klein ist.

Um brauchbare Aussagen über die Eigenschaften einer definierten Grundgesamtheit machen zu können, ist es aus der Sicht des Statistikers nicht nötig, sämtliche zur Gesamtheit gehörenden Elemente zu untersuchen. Vielmehr begnügt man sich mit der Ziehung einer Stichprobe. Wird die Stichprobe nach bestimmten Regeln gezogen, kann man mit Hilfe des statistischen Schlusses von der Stichprobe auf die Grundgesamtheit zu sinnvollen Aussagen über die Beschaffenheit der Grundgesamtheit gelangen. Das Ziehen von Stichproben weist gegenüber Totalerhebungen eine Reihe von generellen Vorteilen auf:

- Stichproben sind wirtschaftlicher als Totalerhebungen.
- Sie liefern trotz des immer auftretenden Stichprobenfehlers oft genauere Ergebnisse (bedingt durch eine bessere Kontrolle, präzisere und umfassendere Datenerhebung).
- Stichprobenverfahren müssen immer dann eingesetzt werden, wenn eine Totalerhebung aus theoretischen Gründen (unendliche Grundgesamtheit) nicht durchgeführt werden kann.

Ist die Entscheidung für die Durchführung einer Stichprobenerhebung gefallen, so muß zunächst ein Stichprobenplan erstellt werden. Ein solcher Plan sollte dem Postulat der Wirtschaftlichkeit entsprechen: Bei vorgegebener Güte der Ergebnisse sollten die Gesamtkosten der Erhebung und Datenaufbereitung minimiert oder bei gegebenem Kostenbudget die Güte der Ergebnisse maximiert werden. Letzteres bedeutet eine Minimierung des Stichprobenfehlers. Die Kosten einer Stichprobe lassen sich relativ einfach durch die Unterscheidung zwischen fixen und variablen (d.h. vom Stichprobenumfang abhängigen) Kosten quantifizieren.

Durch die Fragestellung einer Untersuchung sind u.a. die theoretische und die empirische Grundgesamtheit, die Art der Erhebungs- und Untersuchungseinheiten und die Untersuchungsmerkmale mit dem Problem der Abgrenzung bereits vorgegeben. Zur endgültigen Festlegung des Stichprobenplans gehört eine Entscheidung, die entweder die erforderliche Genauigkeit, die Kosten oder den Stichprobenumfang betrifft. Die Interdependenzen zwischen den verschiedenen Größen der Stichprobenplanung sind in der Abbildung 2.1 zusammenfassend dargestellt.

Im allgemeinen besteht die Zielsetzung in der Bildung einer Stichprobe, die ein 'mikroskopisches Abbild' der Grundgesamtheit liefert, d.h. für die eine

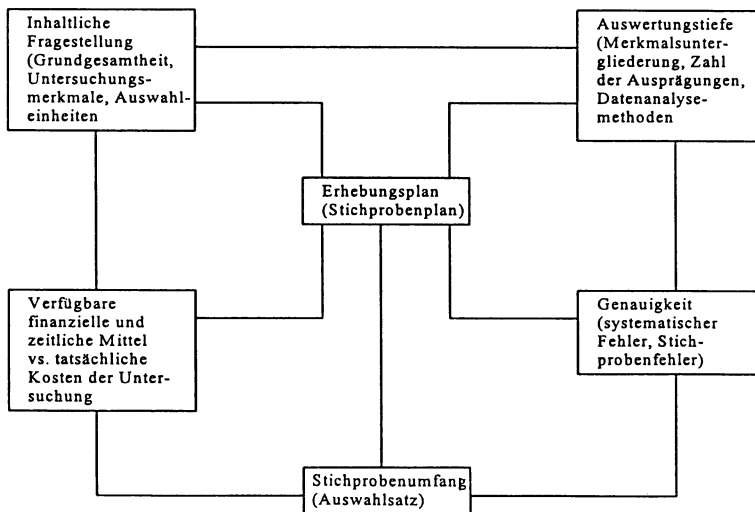


Abb. 2.1: Beziehungen zwischen den Bestimmungsgrößen der Stichprobenplanung
 Quelle: Szameitat, K./Koller S., 1958: Über den Umfang und die Genauigkeit von Stichproben. in: Wirtschaft und Statistik 10, S. 10.

(innerhalb berechenbarer zufälliger Abweichungen) Übereinstimmung der Ausprägungen aller Merkmale (und ihrer Kombinationen) für Grundgesamtheit und Stichprobe besteht. Aufgrund dieser Übereinstimmung genügt die Analyse der zahlenmäßig kleineren Stichprobe für Aussagen über die Grundgesamtheit – unter Nutzung der Zeit- und Kostenvorteile. Das Problem ist nun, Auswahlverfahren zu bestimmen, die eine repräsentative, verkleinerte Abbildung der Grundgesamtheit hervorbringen.

Wenn die Stichprobenelemente nach kontrollierten Zufallsprinzipien aus einer genau definierten Grundgesamtheit ausgewählt worden sind, dann kann man aufgrund der empirischen Befunde in der Stichprobe im Rahmen angegebener Fehlergrenzen auf die Verhältnisse in der Grundgesamtheit schließen. Zufällig heißt: Jedes Element der Grundgesamtheit hat die gleiche Chance (d.h. 'Wahrscheinlichkeit'), für die Stichprobe gezogen zu werden. Zufallsauswahlen werden daher auch als Wahrscheinlichkeitsauswahlen bezeichnet. Die Verknüpfung von Stichprobendaten mit Aussagen über die Grundgesamtheit kann zwei unterschiedliche Zielsetzungen verfolgen:

- **Repräsentationsschluß:** Die Ergebnisse aus der Stichprobe lassen sich auf die Grundgesamtheit verallgemeinern.
- **Inklusionsschluß:** Es sollen formulierte Hypothesen über Grundgesamtheiten anhand von Stichprobenresultaten überprüft werden.

Diese Zielsetzungen lassen sich nur verfolgen, wenn die Stichprobe für die Grundgesamtheit repräsentativ ist, d.h. wenn die Stichprobe als verkleinertes Abbild der Grundgesamtheit angesehen werden kann (Kongruenz zwischen theoretisch definierter Grundgesamtheit und Stichprobe). Die Repräsentativität einer Stichprobe hängt von dem Stichprobenauswahlverfahren und von der korrekten technischen Durchführung ab. Darüberhinaus muß die Grundgesamtheit angebbar und empirisch definierbar sein, damit festgestellt werden kann, welche Grundgesamtheit die Stichprobe abbildet. Von Bedeutung für den Begriff der Repräsentativität ist ferner die Unterscheidung zwischen der eigentlichen angestrebten Grundgesamtheit und der *Erhebungsgrundgesamtheit*. Diese ist entweder

- die zum Zeitpunkt der Auswahl prinzipiell erreichbare Gesamtheit der Untersuchung- bzw. Erhebungseinheiten, falls das Auswahlverfahren direkt auf die Untersuchungseinheiten gerichtet ist,
- oder die tatsächlich repräsentierte Grundgesamtheit, falls sich das Auswahlverfahren auf die symbolische Repräsentation (z.B. durch Lose in einer Urne) der angestrebten Grundgesamtheit stützt.

Letzteres ist von besonderer Bedeutung, denn die eigentliche Auswahl im physikalischen Sinne ist in der Praxis oft ein Ersatzprozeß, da die Stichprobenelemente nicht aus der Grundgesamtheit gezogen werden, sondern aus der Auswahlgrundlage. Dies können z.B. Karteien, Listen, Straßen- und Ortsverzeichnisse sein. Die Identität zwischen theoretischer Grundgesamtheit und Erhebungsgrundgesamtheit ist anzustreben, aber in der Praxis wohl kaum exakt zu realisieren. Es ist zumindest für jede empirische Untersuchung zu überprüfen, ob

- die Auswahlgrundlage vollständig ist, d.h. sämtliche Elemente der Grundgesamtheit enthält,
- nur solche Elemente in die Auswahlgrundlage einbezogen werden, die zugleich der Grundgesamtheit angehören,
- die Möglichkeit besteht, daß ein Element der Grundgesamtheit mehrmals in der Auswahlgrundlage vorkommt.

Jedes noch so korrekte Auswahlverfahren kann schließlich zu einer repräsentativen Stichprobe lediglich im Hinblick auf die Erhebungsgrundgesamtheit führen, nicht jedoch im Hinblick auf die theoretische, d.h. angestrebte Grundgesamtheit.

Die Festlegung des Stichprobenumfangs stellt neben der Wahl des Auswahlverfahrens die wichtigste Entscheidung im Verlauf des Auswahlprozesses dar. Zum einen wird dadurch die Verallgemeinerbarkeit der Stichprobenergebnisse auf die Grundgesamtheit beeinflußt, zum anderen legt der Stichprobenumfang auch die Grenzen der Unterteilbarkeit der Stichprobe in Untergruppen nach den Ausprägungen ausgewählter Merkmale fest. Schließlich ist mit der Festlegung

des Stichprobenumfangs der mit einer Untersuchung verbundene Arbeitsaufwand zu einem beträchtlichen Teil determiniert. Die Entscheidung über den Umfang der Stichprobe sollte aber auch deswegen gründlich überlegt werden, weil Fehler bei der Festlegung des Stichprobenumfangs nachträglich kaum noch korrigiert werden können. Eine nachträgliche Erweiterung der Stichprobe ist oft nur unter erheblichem Mehraufwand möglich, eine zu groß gewählte Stichprobe stellt eine unökonomische Datenbasis dar.

Allgemein gilt das »Gesetz der großen Zahl«. Danach nähern sich die Eigenschaften der Stichprobe mit wachsendem Stichprobenumfang den Eigenschaften der Grundgesamtheit an. Daher sollten möglichst große Stichproben gewählt werden. Andererseits möchte man aus Zeit- und/oder Kostengründen den Stichprobenumfang möglichst gering halten. Eine allgemein verbindliche Untergrenze für den notwendigen Stichprobenumfang gibt es nicht. Es lassen sich lediglich einige Anhaltspunkte für die Bestimmung des Stichprobenumfangs angeben:

- Je stärker sich die Werte der untersuchten Variablen unterscheiden, desto größer sollte der Stichprobenumfang sein.
- Stichprobenumfänge von weniger als 30 Untersuchungseinheiten gelten allgemein als zu klein, um repräsentativ zu sein.
- Die Stichprobengröße muß in Relation zur Größe und Art der Grundgesamtheit gesehen werden, die die Stichprobe abbilden soll.
- Für die Repräsentativität einer Stichprobe ist weniger der relative Anteil der Stichprobengröße an der Größe der Grundgesamtheit von Bedeutung, sondern vielmehr die absolute Größe der Stichprobe.
- Eine wichtige Entscheidungsgröße ist ferner die beabsichtigte Stärke der Aufgliederung der Stichprobe nach ausgewählten Untersuchungsmerkmalen. Die Analyse der nach den Ausprägungen eines Merkmals entstehenden Untergruppen setzt einen Stichprobenumfang voraus, der auch für diese Subpopulationen statistisch abgesicherte Aussagen erlaubt. Die Stärke der Untergliederung wird zum einen von der Zahl gleichzeitig eingeführter Merkmale zur Bildung von Untergruppen bestimmt, zum zweiten von der Zahl der Ausprägungen.

Die wichtigste Entscheidung bei der Planung einer Stichprobe betrifft die Frage, auf welche Art und Weise die Auswahlseinheiten aus der Auswahlgrundlage zu entnehmen sind. Unter Stichprobenverfahren versteht man Methoden, mit deren Hilfe die einzelnen Stichprobenelemente (die Auswahlseinheiten) aus der Grundgesamtheit gezogen werden. Es ist zu entscheiden, ob die Auswahl nach dem Zufallsprinzip, nach subjektiven Überlegungen (bewußte Auswahl) oder willkürlich erfolgen soll. Während die Auswahlverfahren den groben Rahmen für die Stichprobenentnahme abdecken, legen die Auswahltechniken die Anleitungen zur Bestimmung der aus der Auswahlgrundlage zu entnehmenden Elemente fest. Zu unterscheiden ist zwischen der uneingeschränkten Zu-

fallsauswahl, der Anwendung von Ersatzverfahren und der systematischen Zufallsauswahl.

Bei einer willkürlichen Auswahl wird die Ziehung der Elemente aus der Grundgesamtheit nicht durch einen Auswahlplan kontrolliert. Im Gegensatz zu willkürlichen Auswahlen werden bewußte Auswahlen planvoll und aufgrund vorheriger Überlegungen gezielt vorgenommen. Die Auswahl erfolgt nach den Kriterien, die dem Forscher im Rahmen der Fragestellung seiner Untersuchung sinnvoll erscheinen. Für Zufallsauswahlen gilt, daß innerhalb berechenbarer Fehlergrenzen Repräsentativität für alle Elemente, deren Merkmale und Merkmalsausprägungen sichergestellt werden kann, ohne daß Kenntnisse über die Struktur der Grundgesamtheit und die Verteilung der Merkmale vorhanden sein müssen (obwohl es von Nutzen sein kann, wenn diese Informationen vorliegen). Dies wird dadurch gewährleistet, daß jede Auswahleinheit zufällig aus der Erhebungsgrundgesamtheit entnommen wird, so daß für alle Elemente eine gleiche Chance besteht, in die Stichprobe genommen zu werden.

2.2 Zufallsgesteuerte Auswahlen

Nur zufallsgesteuerte Auswahlverfahren gewährleisten die Repräsentativität einer Stichprobe. Bei hinreichend großem Stichprobenumfang kann angenommen werden, daß alle möglichen Merkmale mehr oder weniger entsprechend ihrer Häufigkeit in der Grundgesamtheit auch in der Stichprobe vertreten sind. Untersuchungseinheiten mit Eigenschaften, die häufig in der Grundgesamtheit vorkommen, werden auch in der Stichprobe öfter erfaßt als Einheiten mit Eigenschaften, die in der Grundgesamtheit nur selten vertreten sind. Entsprechendes gilt für Eigenschaftskombinationen. Solange die Auswahl nicht in irgendeiner Art und Weise verzerrt ist, kann davon ausgegangen werden, daß bei hinreichend großem Stichprobenumfang sämtliche Merkmale im Rahmen gewisser Schwankungen entsprechend ihrer Häufigkeit in der Grundgesamtheit auch in der Stichprobe vertreten sind.

Für Zufallsauswahlen gilt das sog. *Gesetz der großen Zahl* von Cournot (1843), dessen Formulierung auf Wahrscheinlichkeitsüberlegungen über das Auftreten einzelner Ereignisse beruht:

- Beim Ziehen einer Zufallsstichprobe wird es sehr selten vorkommen, daß eine Einheit mit einer Merkmalsausprägung gezogen wird, die in der Grundgesamtheit nur selten vorkommt (= Aussage über die Wahrscheinlichkeit des Auftretens einzelner Ereignisse).
- Die Wahrscheinlichkeit, daß der Wert einer statistischen Maßzahl in der Stichprobe beträchtlich (d.h. um mehr als einen vorgegebenen Betrag) von dem entsprechenden Parameter der Grundgesamtheit abweicht, wird umso geringer, je größer die Stichprobe ist (= Wahrscheinlichkeitsaussage über Eigenschaften von Stichproben).

Zufallsgesteuerte Stichprobenverfahren erlauben schließlich eine auf der Wahrscheinlichkeitstheorie beruhende Berechnung sogenannter Fehlerspielräume (sog. »Mutungsbereiche«, »Vertrauensintervalle« oder »Konfidenzintervalle«). Genauer gesagt ist es möglich, einen Bereich zu berechnen, in dem der »wahre Wert« (z.B. eine Maßzahl wie der Mittelwert) der Grundgesamtheit innerhalb einer (selbst bestimmbar) Fehlerwahrscheinlichkeit liegt². Die Abbildung 2.2 illustriert das Verhältnis von Grundgesamtheit und Stichprobe anhand des Rückschlusses von Stichprobenergebnissen auf die Grundgesamtheit.

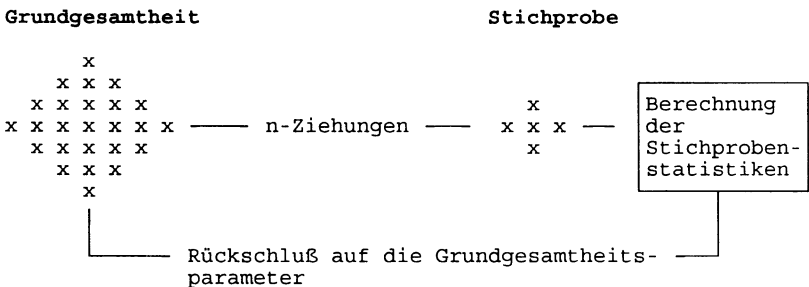


Abb. 2.2: Grundgesamtheit und Stichprobe

Die Schätzung auf der Grundlage einer Stichprobe stellt zwar die bestmögliche Schätzung des Parameters der Grundgesamtheit dar (z.B. der Mittelwert eines Merkmals), die Schätzung ist allerdings mit einem zufälligen Fehler behaftet. Denn würde man nicht nur eine Stichprobe ziehen, sondern eine große Reihe unabhängiger Stichproben, so würden die Mittelwerte dieser unterschiedlichen Stichproben aufgrund rein zufälliger Abweichungen um einen bestimmten »wahren« Grundgesamtheitswert variieren. Aus der Wahrscheinlichkeitstheorie folgt aber, daß wir nur sehr wenige empirische Mittelwerte sehr weit unter bzw. über dem »wahren« Parameterwert der Grundgesamtheit beobachten würden; die meisten Stichprobenmittelwerte würden in der Nähe des »wahren« Mittelwertes der Grundgesamtheit liegen. Wegen dieser Eigenschaft von Zufallsstichproben ist es möglich, mit angebbarer Sicherheit (= Wahrscheinlichkeit) ein (Konfidenz-) Intervall anzugeben, in dem sich ein Stichprobenmittelwert eines Merkmals in der Grundgesamtheit befindet.

In der Forschungspraxis ist nicht immer leicht sicherzustellen, daß die Stichprobeneinheiten wirklich zufällig aus der Menge aller Einheiten der definierten Grundgesamtheit ausgewählt werden. Eine idealtypische Vorgehensweise zur Bildung einer reinen Zufallsstichprobe, die stark von realen Verhältnissen abstrahiert, ist das sog. Lotterieprinzip, das auch als Urnenmodell bezeichnet

² Für die Stichprobe berechnete Maßzahlen (z.B. ein Mittelwert) werden im allgemeinen als »Statistiken« bezeichnet, die entsprechenden Maßzahlen für die Grundgesamtheit heißen »Parameter«.

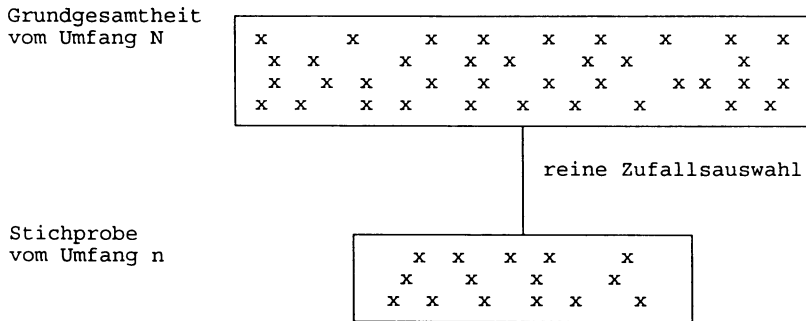


Abb.2.3: Vereinfachtes Urnenschema

wird. Beträgt die Anzahl der Einheiten in der Grundgesamtheit N , der Umfang der Stichprobe n , so kann man sich die einfachste Auswahlvorschrift durch das in Abbildung 2.3 dargestellte Schema veranschaulichen. Die praktische Vorgehensweise läßt sich zusammenfassend wie folgt charakterisieren:

- Im idealen Falle einer Stichprobenziehung ist eine vollständige, durchnummerierte Liste aller zur Grundgesamtheit gehörenden Elemente verfügbar. Die N Elemente der Grundgesamtheit werden durch N in einer Urne befindliche, ebenfalls durchnummerierte Kugeln repräsentiert (d.h. jedem Element der Grundgesamtheit ist eindeutig eine Kugel zugeordnet).
- Im Urnenmodell kann nach gründlichem 'Mischen' von einer ungeordneten, zufällig verteilten Grundgesamtheit ausgegangen werden. Liegt eine solche ungeordnete Grundgesamtheit vor, führt jede Art der Entnahme zu einer streng zufälligen Auswahl (echte Zufallsauswahl). Jedes Element in der Urne hat die gleiche Chance, in die Stichprobe zu gelangen.
- Anschließend werden die Kugeln zufällig Zug um Zug (ohne Zurücklegung) oder jeweils einzeln (mit anschließendem Zurücklegen) und erneutem Mischen nach jeder Ziehung aus der Urne entnommen.

Man nennt $f = n/N$ den Auswahlatz; er gibt an, welcher Teil der Gesamtheit in die Stichprobe gelangt. Jeder Auswahlvorgang setzt das Vorhandensein einer Auswahlgrundlage voraus, in der jede Einheit in irgendeiner Form registriert ist. Auswahlgrundlage können ungeordnete Listen, Verzeichnisse, Karteien etc. sein. Theoretisch wäre es sehr einfach, für jede Einheit einen Zettel anzulegen mit dem Namen, der Nummer oder einem ähnlichen Identifikationskriterium, die Zettel in eine Urne zu legen und die gewünschte Anzahl von Einheiten nach gründlichem Durchmischen zufällig zu entnehmen.

Wenn die Grundgesamtheit sehr groß ist, kann eine zufällige Auswahl auf diese Weise aufgrund des zu großen Aufwandes nicht mehr durchgeführt werden. Es ist aber möglich, das Prinzip der Urnenziehung auch bei umfangreichen Grundgesamtheiten beizubehalten. Der Kunstgriff besteht darin, daß man je-

dem Element der (ungeordneten) Grundgesamtheit eine Zahl zuordnet, wobei sich die Nummern nicht wiederholen dürfen. In die Urne legt man zunächst 10 Kugeln, die mit den Ziffern des Dezimalsystems gekennzeichnet sind. Mit diesen 10 Kugeln können nun beliebig viele Elemente der Grundgesamtheit gezogen werden. Ist die Anzahl der Elemente in der Stichprobe eine mehrstellige Zahl, so kann man diese Zahl zusammensetzen (Einer, Zehner, Hunderter usw.). Die einzelnen Ziffern dieser mehrstelligen Zahl werden dann getrennt und unabhängig voneinander (mit Zurücklegen) aus der Urne mit den 10 Kugeln gezogen.

Eine Zufallsauswahl nach dem Urnenmodell ist in der Praxis nur selten durchführbar. Es sind deshalb modifizierte Verfahren zu suchen, die grundsätzlich das Vorgehen beim Urnenmodell beibehalten, aber doch nicht an ideale Ausgangsbedingungen gebunden sind. Sie müssen die Erfordernisse des Urnenmodells mit den praktischen Gegebenheiten in Übereinstimmung bringen. Das Vorhandensein dieser Übereinstimmung stellt sich nicht automatisch ein, sondern muß von Fall zu Fall geprüft werden und zur Wahl des entsprechend geeigneten Auswahlverfahrens führen. Die modifizierten Auswahlverfahren beinhalten ebenfalls prinzipielle Regeln zur Konstruktion von Zufallsstichproben. In Abhängigkeit von der Spezifik der Einheiten und der Art der Auswahlgrundlage sind die in Abbildung 2.4 dargestellten Verfahren zur zufälligen Auswahl der Stichprobenelemente anwendbar.

(1) Auswahl nach Zufallszahlen:

Da man in der Praxis nicht überall Urnen zur Hand hat, verwendet man sog. *Zufallszahlentabellen*. Sie enthalten eine große Anzahl von mehrziffrigen Zahlen, die man als »Urnen auf Vorrat« bezeichnen kann. In der Praxis wird daher die echte Zufallsauswahl in der Regel durch bereits vollzogene Zufallszahlengenerierung ersetzt³. Es ist dann nur noch notwendig, dieser Tabelle die notwendigen Zahlen zu entnehmen.

Um die Auswahlseinheiten aus der Auswahlgrundlage mittels Zufallszahlen entnehmen zu können, muß man zunächst sämtliche Einheiten mit einer Ordnungsnummer versehen. Die Verwendung von Zufallszahlen ist dann sehr einfach. Eine Stichprobe der Größe n erfordert n Zufallszahlen. Sollen z.B. aus einer Gesamtheit von $N = 100$ Einheiten insgesamt $n=10$ Elemente zufällig ausgewählt werden, so wählt man zunächst eine beliebige Spalte und Zeile der Zufallstafel aus und beginnt vertikal oder horizontal fortlaufend die Zufallszahlen auszuwerten. Die Tabelle 2.1 zeigt einen Ausschnitt aus einer Zufallstafel. Für das beschriebene Beispiel werden jeweils die beiden letzten Ziffern

³ Zufallszahlen werden heute mit Hilfe der EDV durch Simulationsverfahren erzeugt (Zufallszahlengenerator). Nahezu jedes Statistiklehrbuch enthält Auszüge aus Zufallszahlentabellen (vgl. z.B. Bortz, J., 1984: Lehrbuch der empirischen Sozialforschung, Berlin, S. 543).

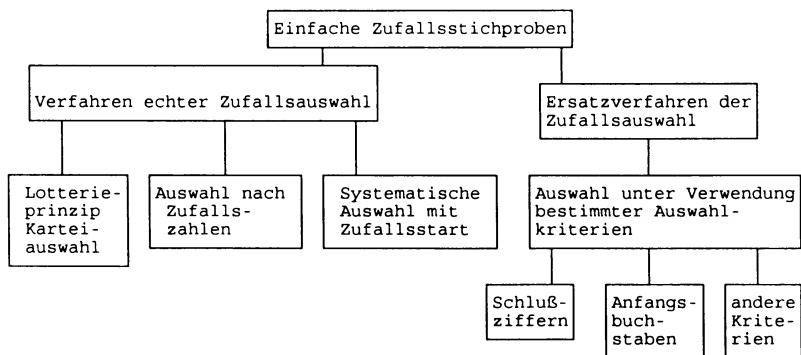


Abb. 2.4: Klassifikation einfacher Techniken der Zufallsauswahl

Tab. 2.1: Zufallszahlen

441425	157136	711327	163069	797674	956480
926356	401559	258751	008863	665288	358055
862351	326815	293872	942479	424573	053784
320012	939223	461432	820309	191387	278381
035136	942594	307986	326250	541309	058683
204900	633642	531353	376165	879052	386383
318010	809195	359393	290413	674787	931928
675870	814221	556217	724198	067764	329585
054487	329788	906335	284258	656254	516388
368934	040032	960883	359816	301410	011515
758456	572604	299904	426646	152043	922405

der in der Tabelle enthaltenen Zufallszahlen zeilenweise abgelesen. Begonnen wird in Zeile 5 und Spalte 2. Dann ergibt sich:

94, 86, 50, 09, 83, 00, 42, 53, 65, 52.

Diese Zahlen sind dann die zufällig ausgewählten Identifikationsnummern für die Ziehung der nummerierten Einheiten der Grundgesamtheit. Anschließend werden die zu den gezogenen Ziffern gehörenden Einheiten einer Kartei oder Liste herausgesucht und dadurch die Elemente der Stichprobe bestimmt. Bei jeder Auswertung ist zu beachten, daß das erste Element der Grundgesamtheit die Nummer Null (00) hat, das zweite die Nummer Eins (01), usw. Diese Verfahren sind immer dann noch relativ aufwendig, wenn nicht die Kartei selbst schon durchnummeriert in eine EDV-Anlage gespeichert ist. Im Falle von sehr großen Grundgesamtheiten empfiehlt es sich, Modifikationen der reinen Zufallsauswahl zu verwenden.

Die Verwendung von Zufallszahlen vereinfacht zwar den Auswahlvorgang bereits beträchtlich, ist aber bei umfangreichen Stichproben immer noch sehr arbeitsaufwendig, weil sämtliche Elemente der Grundgesamtheit verfügbar gemacht und aufgelistet werden müssen (falls keine EDV-gespeicherte Liste existiert). Aus pragmatischen Gründen weicht man deshalb von der Realisierung des Urnenmodells als echte Zufallsauswahl häufig ab, indem man sie durch verschiedene, praktikablere Ersatzverfahren weiter vereinfacht. Die folgenden Auswahlverfahren sind Modifikationen der reinen Zufallsauswahl.

(2) Ersatzverfahren für echte Zufallsauswahlen:

Bei einer *systematischen Zufallsauswahl* wird nur die erste zu ziehende Einheit (die erste zu ziehende Karte einer Kartei oder die erste zu berücksichtigende Einheit in einer Liste) rein zufällig bestimmt, etwa durch Herausgreifen einer Zufallszahl aus einer entsprechenden Tabelle (= Zufallsstart). Ab dieser Einheit wird anschließend jedes x -te Element ausgewählt, bis n Stichprobenelemente zusammen sind. Der konstante Abstand » x « heißt *Zählabstand*. Vielfach wird das Entnahmeintervall $x = N/n$ gesetzt (Umfang der Grundgesamtheit dividiert durch den angestrebten Stichprobenumfang). Durch eine systematische Auswahl spart man sich den aufwendigen Lotterievorgang (Durchmischen und zufälliges Ziehen von Elementen aus einer Liste oder Kartei).

Als einen Sonderfall der systematischen Auswahl kann man ein Verfahren ansehen, das den Umfang der Grundgesamtheit in irgendeiner Weise durch ein Längenmaß ausdrückt, etwa durch die Länge des Karteikastenstapels, wenn jede Einheit durch eine Karteikarte repräsentiert ist. Angenommen, die Gesamtheit ist x cm lang. Die Karten, die man bei geeignetem Zufallsstart im Abstand x/n cm entnimmt, stellen dann die Stichprobeneinheiten dar.

Voraussetzung für diese Ersatzverfahren ist jedoch, daß eine zufällige Ordnung der Elemente in der Liste oder Kartei gegeben ist. Ein Verstoß gegen das Zufallsprinzip läge z.B. dann vor, wenn die Kartei nach einer bestimmten Variablen systematisch organisiert ist und das Auswahlssystem Elemente des Karteiordnungssystems enthält. Ein Beispiel ist eine Karteiordnung nach den Anfangsbuchstaben von Familiennamen und ein sich ebenfalls darauf stützendes Auswahlssystem (z.B. Buchstabenverfahren). Um diese Gefahr zu vermeiden, muß man die Ordnung in der Erhebungsgrundlage genau prüfen und bei Vorliegen von Regelmäßigkeiten die systematische Entnahme entsprechend modifizieren, etwa indem man das Entnahmeintervall variiert.

Die systematische Auswahl ist eine wesentlich schnellere, einfachere und damit ökonomischere Verfahrensweise. Der technisch unkomplizierte Arbeitsgang kann auch von ungeschultem Personal durchgeführt werden. Neben der systematischen Auswahl mit Zufallsstart wurde eine Reihe weiterer Verfahren entwickelt, die ebenfalls den Auswahlvorgang vereinfachen (unter Beibehaltung des Zufallsprinzips).

Das *Schlußziffernverfahren* hängt noch am engsten – als eine alternative Form der systematischen Zufallsauswahl – mit dem Originalverfahren zusammen. Ausgehend von einer lückenlosen Durchnummerierung der Grundgesamtheitselemente werden entsprechend dem vorgegebenen Stichprobenumfang jene Elemente in die Stichprobe genommen, deren Nummerierung auf eine bestimmte Endziffernfolge endet. Die Endziffern oder Endzifferkombinationen werden per Zufallsauswahl bestimmt. Jeder Auswahlatz n läßt sich durch eine entsprechende Definition der zu erfüllenden Endzifferbedingungen mit dieser Methode realisieren. Sollen zum Beispiel 10% der Grundgesamtheitselemente in die Stichprobe aufgenommen werden, wird als Bedingung für die Aufnahme in die Stichprobe die letzte Ziffer für die zufällig ausgewählte Zahl genommen. Wurde etwa die Zahl »4« ermittelt, so werden alle Elemente in die Stichprobe aufgenommen, die mit der Ziffer »4« enden. Für den Fall eines 20%igen Auswahlatzes müßte eine zweite Zahl gezogen werden (z.B. die »6«). In diesem Fall werden dann sämtliche Elemente in die Stichprobe aufgenommen, die mit einer »4« oder einer »6« enden. Entsprechend lassen sich auch komplexere Auswahlätze bilden. Für eine Auswahlatz von 3% der Grundgesamtheit können z.B. die Elemente mit der Schlußzifferkombination »23«, »46« und »19« berücksichtigt werden.

Handelt es sich bei der zu untersuchenden Gesamtheit um Personen und enthält die Auswahlgrundlage den Namen der Personen, so können bestimmte Anfangsbuchstaben zur Bestimmung der Stichprobeneinheiten dienen (*Buchstabenverfahren*). Die Auswahl nach dem Anfangsbuchstaben kommt einer direkten Zufallsauswahl immer dann nahe, wenn zwischen den zu untersuchenden Merkmalen und dem Namensanfang als Auswahlkriterium keine Zusammenhänge bestehen. Ein Beispiel für einen systematischen Auswahlfehler können ethnische Minderheiten sein, deren Nachname überproportional mit einem bestimmten Anfangsbuchstaben beginnen. Ein weiteres bekanntes Beispiel ist die regionale Ungleichverteilung der Anfangsbuchstaben (= Zusammenhang zwischen REGION und dem Auswahlkriterium NAMENSANFANG). Voraussetzung ist auch, daß mit der Wahl bestimmter Anfangsbuchstaben keine Personengruppe überrepräsentiert ist, die ganz bestimmte Merkmale aufweisen könnte (Verzerrung der Ergebnisse durch den sogenannten Klumpeneffekt)¹. Man wählt daher oft Personen nach mehreren (näher zu bestimmenden) Anfangsbuchstaben aus. Das Problem systematischer Verzerrungen sei an einem Beispiel erläutert:

»Nehmen wir an, daß im Kreis Neuhaus am Rennweg im Bezirk Suhl durch eine Stichprobenuntersuchung der Anteil der Beschäftigten geschätzt werden soll, die in der Glasindustrie beschäftigt sind. In die Auswahl sollen alle berufstätigen Personen mit den Anfangsbuchstaben M und G einbezogen werden.

¹ Vgl. Schach, E. / Schach, S., 1978: Pseudorauswahlverfahren bei Personengesamtheiten I: Namensstichproben, in: Allgemeines Statistisches Archiv 62, S. 379–396.

Nun ist bekannt, daß in einigen Orten dieses Kreises (Lauscha, Neuhaus, Ernsththal mit einer traditionell starken Glasindustrie) unter den Beschäftigten in diesem Industriezweig die Namen Müller und Greiner (historisch bedingt) außerordentlich häufig sind. Unter den Berufstätigen mit dem Namensanfang M und G wird demnach der Anteil der Beschäftigten der Glasindustrie wesentlich über dem entsprechenden Anteil in der Gesamtheit aller Beschäftigten des Kreises liegen. Eine derart durchgeführte Auswahl nach dem Namensanfang wäre also für die Untersuchung der Berufsstruktur im Kreis Neuhaus am Rennweg ungeeignet, da hier ein Zusammenhang zwischen dem Namensanfang und dem Untersuchungsmerkmal BERUF vorliegt« (Schwarz, H., 1975: Stichprobenverfahren, München/Wien, S. 23).

Sind die Elemente einer Gesamtheit Personen und enthält die Auswahlgrundlage deren Geburtsdatum, so können Personen, die an einem bestimmten Tag des Jahres (oder des Monats) geboren sind, als Zufallsstichprobe angesehen werden (*Geburtsstagsverfahren*). Wenn keine Zusammenhänge zwischen dem Geburtstag einer Person und interessierenden Merkmalen bestehen, gehört diese Auswahlvorschrift in jedem Falle zu den zufallsgesteuerten Verfahren.

Erhebungseinheiten müssen nicht unbedingt durch eine Kartei (oder Liste, EDV-Datei) symbolisch repräsentiert werden. Es lassen sich in bestimmten Fällen auch Flächen verwenden (*Flächen- oder allgemeiner Gebietsauswahl*). Sie kann u.a. dann angewendet werden, falls eine Kartei etc. nicht verfügbar ist. Personen lassen sich z.B. auch durch den Ort, an dem sie wohnen, symbolisch repräsentieren. Ereignisse, die registriert werden sollen, lassen sich durch die Orte, an denen sie geschehen, symbolisch repräsentieren. Die Auswahlseinheiten sind dann definierte und abgegrenzte Gebiete (Flächen), auf denen Personen oder Haushalte wohnen oder untersuchungsrelevante Ereignisse stattfinden. Auswahlgrundlage könnten dann etwa Landkarten oder Stadtpläne sein. Die räumlichen Einheiten müssen geeignet sein, die interessierenden Erhebungseinheiten zu bestimmen.

Im Fall einer Gebietsauswahl stellt in der Regel die Flächenstichprobe die erste Stufe einer mehrstufigen Auswahl (ausführlich s. weiter unten) dar (Flächen als Primäreinheiten). Aus den Flächenstichproben wird häufig in einem zweiten Schritt per Lotterieverfahren oder unter Verwendung von Zufallszahlen nochmals eine Auswahl entnommen, die eigentlichen Stichprobenelemente (etwa Personen, Haushalte oder Ereignisse als Sekundäreinheiten).

Der Hauptvorteil einer Flächenstichprobe gegenüber den Formen der Karteiauswahl besteht darin, daß eine Abgrenzung der Auswahlseinheiten und deren eindeutige, fortlaufende Nummerierung vor der eigentlichen Auswahl nicht notwendig ist. Sie braucht nur für die Elemente zu erfolgen, die Teil der ausgewählten Flächen sind.

Dem Vorteil steht allerdings ein Nachteil gegenüber. Nahezu immer enthalten Primäreinheiten (die Flächen) unterschiedlich viele Elemente. »Dies muß bei der Auswahl der Primäreinheiten berücksichtigt werden, da die Elemente

aus größeren Primäreinheiten eine niedrigere Auswahlwahrscheinlichkeit besitzen, falls alle Primäreinheiten mit gleicher Wahrscheinlichkeit gezogen werden. Eine Möglichkeit trotz ungleicher Größe der Primäreinheiten die gleichen Auswahlwahrscheinlichkeiten für jedes Element der Grundgesamtheit zu erhalten, besteht darin, Auswahlwahrscheinlichkeiten für die Primäreinheiten zu verwenden, die proportional zur Größe der Primäreinheiten sind, und aus jeder Primäreinheit dieselbe Zahl von Sekundäreinheiten zu ziehen« (PPS-Design, »Probability Proportional Size«; Schnell, R./Hill, P.B./Esser, E., 1988: Methoden der empirischen Sozialforschung, München/Wien, S. 266. Ausführlich: Sudman, S., 1976: Applied sampling, New York, S. 131–170).

Verzerrungen lassen sich auch vermeiden, indem man die Grundgesamtheit zunächst in gleich große Flächen zerlegt (d.h. schichtet) und erst aus diesen Schichten eine Flächenstichprobe zieht (ausführlich s. »geschichtete Zufallsauswahlen« im folgenden Unterpunkt). Einfache Zufallsauswahlen weisen generell einige Nachteile auf:

- Häufig ist die betrachtete Population so groß, daß es technisch gar nicht möglich ist, eine einfache Auswahl zu entnehmen.
- Bei regional sehr weit gestreuten Populationen ist das Verfahren aufwendig, da die zufällig ausgewählten Erhebungseinheiten weit über das Gebiet verteilt sind.
- Eine weitere Schwierigkeit kann sich daraus ergeben, daß die für die Auswahl notwendige Durchnummerierung (technisch) nicht vorgenommen werden kann.
- Interessante Untergruppen der Population, die zahlenmäßig nur gering vertreten sind, werden auch in der Zufallsstichprobe mit sehr wenigen Untersuchungseinheiten vertreten sein, so daß sie nicht gesondert ausgewertet werden zu können.

In solchen Fällen ist es ratsam, das Auswahlverfahren so zu modifizieren, daß es den praktischen Anforderungen eher gerecht wird. Um die genannten Nachteile zu vermeiden, wurden die Verfahren der geschichteten, der mehrstufigen und der Klumpen-Stichprobe entwickelt. Für diese komplexen Zufallsauswahlverfahren wird nicht mehr die Forderung aufrecht erhalten, jede Einheit der Grundgesamtheit müsse die gleiche Chance haben, um in die Auswahl zu gelangen.

2.3 Geschichtete Zufallsauswahl

Bei der geschichteten Stichprobe zieht man aus geeignet gewählten Teilgesamtheiten, aus denen sich eine Grundgesamtheit zusammensetzt, jeweils eine einfache Zufallsstichprobe. Sinnvoll und möglich ist das Ziehen von geschichteten Stichproben immer dann, wenn eine Grundgesamtheit in Schichten oder Klassen zerfällt oder zerlegbar ist, die jede für sich möglichst homogen ist. Die

Schichtung ist eine Gliederung der N Auswahleinheiten einer Gesamtheit in allgemein k sich gegenseitig ausschließende (d.h. disjunkte) Teilgesamtheiten vom Umfang N_i ($i = 1, \dots, k$), die als Schichten (»strata«) bezeichnet werden. Jede Auswahleinheit wird genau einer Schicht zugeordnet. Analog zur einfachen Zufallsstichprobe wird nun jede Schicht wie eine eigene Grundgesamtheit behandelt, aus der eine Stichprobe von noch festzulegendem Umfang n_i gezogen wird. Um das Wesen der geschichteten Auswahl verständlich zu machen, kann man auf das bereits vorgestellte Urnenmodell zurückkommen. Bei der einfachen Zufallsauswahl ging man von einer Urne aus, in der die Kugeln als symbolische Repräsentationen der Erhebungseinheiten gut durchgemischt waren, bevor gezogen wurde. Das Modell der geschichteten Auswahl führt im Prinzip zu einem 'parallelen Urnenschema': Die Elemente der Grundgesamtheit (bzw. die entsprechenden Kugeln) sind auf zwei oder mehrere Urnen verteilt, d.h. es wird nicht mehr nur eine Auswahl getroffen, sondern insgesamt L separate Zufallsauswahlen aus L separaten Urnen (L = Anzahl der Ausprägungen des Schichtungsmerkmals), die zusammen die interessierende Grundgesamtheit repräsentieren. Schematisch kann man dieses Vorgehen auch an der Abbildung 2.5 veranschaulichen.

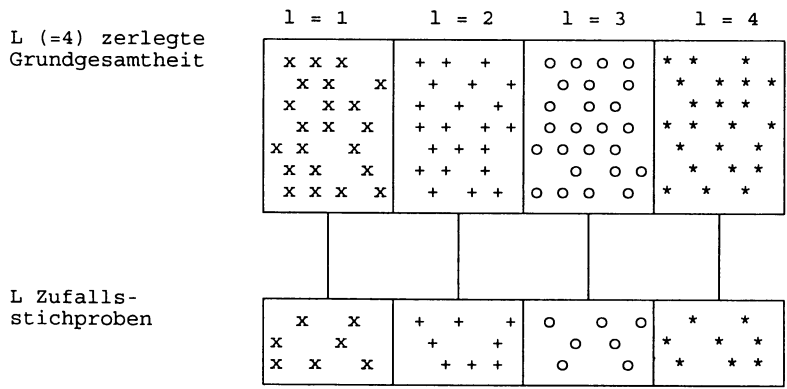


Abb. 2.5: Geschichtete Zufallsstichprobe als 'paralleles Urnenschema'

Die Schichtung orientiert sich an den Ausprägungen eines auszuwählenden (Untersuchungs-) Merkmals. An erster Stelle steht somit die Frage, welches Merkmal der Schichtung zugrundegelegt werden soll. Abgesehen von dem einfachen Fall nur eines einzigen Schichtungsmerkmals erhebt sich ferner die Frage, ob es möglich, zweckmäßig oder etwa notwendig ist, mehrere Erhebungsmerkmale für eine Schichtung der Gesamtheit zu verwenden. Aus praktischen Erwägungen sollte die Zahl der Schichtungsmerkmale möglichst klein gehalten werden, damit nicht zu viele (durch Merkmalskombinationen definierte) Schichten unterschieden werden müssen und die Auswahltechnik

nicht zu unübersichtlich und fehleranfällig wird. Diese Gründe sprechen prinzipiell für nur ein einziges Schichtungsmerkmal.

- Für die Schichtung eignet sich vorrangig ein solches Merkmal, das in der Stichprobenerhebung besonders bedeutungsvoll ist und an das relativ hohe Genauigkeitsanforderungen oder -erwartungen geknüpft werden.
- Für die Schichtung wird das Merkmal gewählt, das mit einem (mehreren) bedeutungsvollen Untersuchungsmerkmal(en) erwiesenermaßen oder hypothetisch besonders hoch korreliert ist.
- Ausgangspunkt für die Konstruktion einer geschichteten Auswahl kann auch die Absicht sein, die Stichprobendaten getrennt für bestimmte Gruppen (Schichten) von Untersuchungseinheiten auszuwerten. In solchen Fällen ist es sinnvoll, schon bei der Auswahl sicherzustellen, daß diese Gruppen (bzw. Teilpopulationen) in der Stichprobe zuverlässig repräsentiert sind und die für die Anwendung von speziellen Auswertungstechniken erforderlichen Mindestfallzahlen in den einzelnen Gruppen gegeben sind.

Bei der geschichteten Stichprobe ist der gesamte Stichprobenumfang n auf die einzelnen Schichten zu verteilen. Abgesehen von der nicht üblichen gleichmäßigen Aufteilung des gesamten Stichprobenumfangs auf die einzelnen Schichten, sind hier drei Hauptformen der geschichteten Auswahl zu unterscheiden:

a) Proportional geschichtete Stichprobe: Hier geht man von der Überlegung aus, daß aus großen Schichten in der Grundgesamtheit auch große Stichproben entnommen werden sollten und aus kleinen Schichten entsprechend kleinere Stichproben. Dadurch ordnet man großen Schichten ein größeres Gewicht in der Stichprobe zu als kleineren Schichten. Die Grundformel ist folglich: Jeder Schicht wird zufällig eine Anzahl von Beobachtungseinheiten entnommen, die dem Anteil dieser Schicht an der Grundgesamtheit entspricht. Der Wahlsatz für die allgemein i -te Schicht berechnet sich als $n_i = n(N_i/N)$. Voraussetzung ist natürlich, daß die Verteilung des Schichtungsmerkmals in der Grundgesamtheit bekannt ist. Der Vorteil gegenüber einer einfachen Zufallsauswahl besteht darin, daß die Anteilswerte des für die Auswertung bedeutsamen Schichtungsmerkmals exakt mit den Werten der Grundgesamtheit übereinstimmen.

b) Disproportional geschichtete Auswahl: In diesem Fall entsprechen die Fallzahlen der gruppenspezifischen Zufallsstichproben nicht den Anteilen der Schichten in der Grundgesamtheit. Angestrebt wird eine Aufteilung des Gesamtstichprobenumfangs n , bei der eine oder mehrere Schichten in der Stichprobe stärker oder schwächer berücksichtigt werden und dafür die Wahlsätze in diesen Schichten erhöht bzw. verringert werden. Die Schichtanteile werden gegenüber der Grundgesamtheit somit bewußt verzerrt. Häufig ist es sinnvoll, Untersuchungseinheiten mit nur geringem Anteil in der Grundgesamtheit stärker in der Stichprobe zu repräsentieren als Untersuchungsein-

heiten mit hohem Anteil in der Grundgesamtheit. Für anteilmäßig kleine Gruppen werden so detaillierte Auswertungsmöglichkeiten geschaffen. Für eine gruppenweise getrennte Auswertung der Erhebungsdaten ist eine Mindestanzahl von Untersuchungseinheiten erforderlich, um statistische Kennwerte und Beziehungsmaße berechnen zu können. Bei einer proportionalen Auswahl würde bei kleinen Teilgesamtheiten der Gesamtumfang der Stichprobe sehr stark anwachsen, damit diese Teilgruppen einen für Auswertungen adäquaten Umfang hätten.

Disproportionale Schichtung ist auch immer dann notwendig, wenn aus bestimmten Gründen eine genau gleiche Anzahl von Untersuchungseinheiten je Gruppe angestrebt wird (etwa für spezifisch tabellarische Auswertungen mit einem konstant gehaltenen Leitmerkmal, das im Hinblick auf die angestrebte Aussage der Untersuchung bzw. aufgrund des Untersuchungszieles identisch mit dem Schichtungsmerkmal ist).

Bei vorgegebenem Auswahlumfang ist die »Effektivität« geschichteter Stichproben nahezu immer höher als bei der einfachen Zufallsauswahl, d.h. die Verlässlichkeit von Verallgemeinerungen von der Stichprobe auf die Grundgesamtheit erhöht sich. Dieser Effekt wird als »Schichtungseffekt« bezeichnet. Der Schichtungseffekt ist um so größer, je homogener die Schichten in sich und je heterogener die Schichten untereinander sind (d.h. je kleiner die Streuung innerhalb einer jeden Schicht und je größer die Streuung zwischen den einzelnen Schichten bezüglich der Verteilung untersuchungsrelevanter Merkmale ist). Daraus erwächst die Forderung, als Schichtungsmerkmale diejenigen Merkmale auszuwählen, die mit den eigentlichen Erhebungsmerkmalen möglichst eng zusammenhängen.

2.4 Die Klumpenstichprobe

In der Praxis ist die Datenerhebung einer einfachen Zufallsauswahl oder einer geschichteten Stichprobe häufig mit einem erheblichen Aufwand verbunden. Insbesondere bei Stichproben, die eine sehr große Fläche zugrunde legen, treten mehrere Probleme auf:

- Die Beschaffung einer vollständigen Auswahlliste für die Grundgesamtheit ist – sofern überhaupt möglich – außerordentlich aufwendig.
- Wird aus einer solchen Auswahlliste eine einfache oder geschichtete Zufallsstichprobe gezogen, wird diese geographisch sehr weit streuende Auswahlseinheiten liefern. Dies führt zu hohen Erhebungskosten.

Die Klumpenauswahl (auch *Auswahl von geschlossenen Erfassungsgruppen* genannt) versucht, diese Probleme zu umgehen. Kleinste Einheiten sind nicht mehr die Merkmalsträger selbst, sondern nur Gruppen (»cluster«) von Merkmalsträgern. Diese Klumpen müssen die gleiche Struktur wie die Grundgesamtheit aufweisen, d.h. der einzelne Klumpen sollte wie in einem Mikrokosmos die

Grundgesamtheit widerspiegeln. Oder anders formuliert: Innerhalb der einzelnen Klumpen kann Heterogenität bezüglich der Merkmalsausprägungen der einzelnen Untersuchungseinheiten vorliegen, zwischen den Klumpen herrscht Homogenität, d.h. jeder Klumpen besitzt annähernd die gleiche Struktur. Die Grundgesamtheit wird nach einer natürlichen Gruppierung, die man als *Klumpenbildung* bezeichnet, in k Teilgesamtheiten aufgeteilt, die ihrerseits die Auswahlseinheiten bilden (sog. Primäreinheiten). Nach dem Zufallsprinzip wählt man dann im ersten Schritt einige Klumpen aus und führt innerhalb der Klumpen eine Totalerhebung durch, d.h. auf der zweiten Stufe wird kein eigentlicher Auswahlvorgang mehr vorgenommen. Zur Veranschaulichung der Klumpenauswahl bietet sich wiederum das Urnenmodell an.

- a) Die N Kugeln einer Grundgesamtheit werden zunächst in eine Anzahl von (nicht notwendigerweise gleich umfangreicher) Teilurnen (= Primäreinheiten) aufgeteilt.
- b) Ein Teil der gebildeten Urnen wird per Zufallsauswahl ausgewählt – die Urnen (nicht die Kugeln) bilden hierbei die Stichprobenelemente.
- c) Die Summe (n) aller in den ausgewählten Urnen befindlichen Kugeln bilden die Stichprobe.

Bei der Klumpenauswahl divergieren somit die Auswahlseinheiten (= Klumpen) und die Analyseeinheiten (= sämtliche Elemente der ausgewählten Klumpen). Die Analyseeinheiten werden somit nicht einzeln in die Stichprobe aufgenommen, sondern jeweils in Gruppen der Stichprobe zugeführt. Das setzt zunächst voraus, dass sich die Grundgesamtheit in einfach zu unterscheidende Klumpen zerlegen lässt. Die Abbildung 2.6 veranschaulicht das Grundprinzip der Klumpenauswahl anhand eines einfachen Beispiels.

Das Klumpen-Auswahlverfahren ist immer dann vorteilhaft, wenn eine Liste der Elemente der Grundgesamtheit nicht vorhanden ist, wohl aber eine Liste der Teilkollektive (der zusammengefassten Elemente). Der Hauptvorteil liegt dann in der unkomplizierten und schnelleren Durchführung der Stichprobenbildung. Die Klumpen dürfen allerdings selbst nicht zu groß sein, damit auch sämtliche Elemente der ausgewählten Klumpen die Untersuchungseinheiten bilden können.

Man unterscheidet bei der Klumpenbildung zwischen natürlichen und sogenannten künstlichen Klumpen. Natürliche Klumpen gehen auf bereits vorhandene Gliederungen der Grundgesamtheit zurück, z.B. räumliche Gliederungsmerkmale wie Bezirke, Kreise etc. Künstliche Klumpen entstehen nach Gruppierungen, die vom Forscher eigens zu Analysezwecken gebildet werden. Werden als Primäreinheiten eine Vielzahl von disjunkten Flächenelementen angenommen und von 1 bis k durchnummeriert, im zweiten Schritt v Klumpen zufällig ausgewählt, dann spricht man von einer *Flächenstichprobe* (»area sampling«).

Die spezielle Gefahr der Klumpenauswahl liegt im Auftreten des sogenannten *Klumpeneffektes*. Immer dann, wenn ausgewählte Klumpen in sich homo-

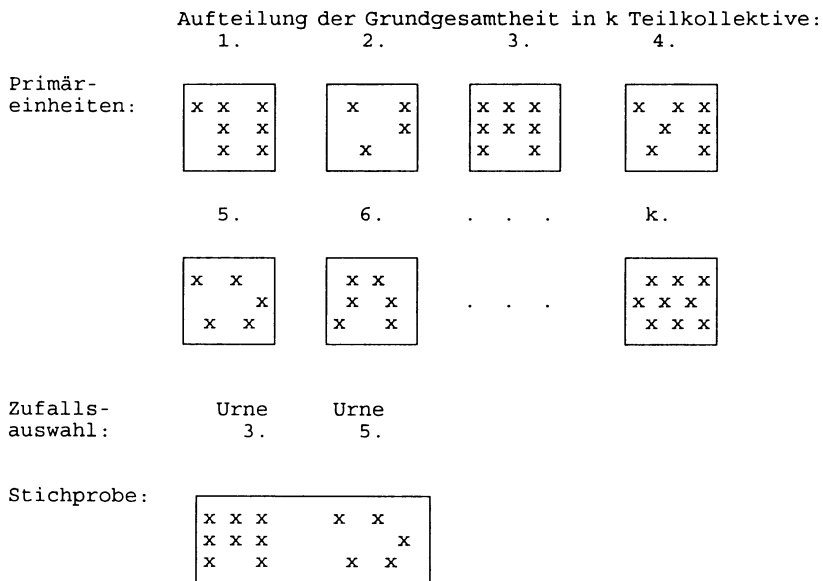


Abb. 2.6: Das Auswahlmodell der Klumpenauswahl

gen, aber von der Grundgesamtheit stark abweichend strukturiert sind, kann es leicht zu gravierenden Ergebnisverzerrungen kommen. Flächenstichproben mit sehr kleinen Segmenten sind z.B. oft dadurch gekennzeichnet, daß die einzelnen Segmente relativ homogen sind, wie etwa ländliche Teilgebiete, Arbeiterstadtbezirke, Villenviertel etc. Ist dagegen die Struktur der einzelnen Klumpen heterogen im Sinne einer guten verkleinerten Abbildung der Grundgesamtheit, dann kann mit wenigen Klumpen eine kostengünstige und darüberhinaus repräsentative Stichprobe gezogen werden. Es läßt sich zeigen, daß die relative Ungenauigkeit einer Klumpenstichprobe gegenüber einer einfachen Zufallsauswahl sowohl von der Homogenität als auch von der Anzahl der Elemente pro Cluster abhängt: Je größer die Homogenität der Elemente innerhalb der Cluster und je größer die Cluster, desto ungenauer sind Klumpenstichproben. Eine große Anzahl zufällig ausgewählter Cluster mit jeweils sehr wenigen Elementen pro Cluster erhöhen dagegen die Genauigkeit der Stichprobe relativ zur Grundgesamtheit. Allerdings steigen bei zunehmender Zahl der ausgewählten Klumpen die Kosten, so daß ein vernünftiger Kompromiß gefunden werden muß. Außerdem ist es nicht immer einfach, möglichst kleine und heterogene Auswahlseinheiten zu bilden, was u.a. an den Informationslücken über die Varianzen wichtiger Merkmale innerhalb einer Auswahlinheit liegt.

Die oben erwähnte Verwendung der Buchstabenauswahl ist auch eine Klumpenauswahl. Ist die Grundgesamtheit regellos geordnet, entspricht die Klum-

penauswahl einer reinen Zufallsauswahl. Ansonsten wird das Bild von der Grundgesamtheit in der Stichprobe durch den negativen Anordnungseffekt oder Klumpeneffekt verzerrt. Wählt man z.B. aus einer nach Namen geordneten Kartei alle Personen mit dem Anfangsbuchstaben Z aus, so werden dort wahrscheinlich Heimatvertriebene und Flüchtlinge aus den ehemaligen deutschen Ostgebieten sowie ihre Nachkommen überrepräsentiert sein (d.h. polnische oder slawische Namen).

Bei sehr großen Klumpen (z.B. bei Gebietsauswahlen: Regionen, Städte, Stadtteile, Gemeinden oder Stimmbezirke etc.) wird das Klumpenverfahren zu- meist mit anderen Auswahlverfahren kombiniert, so daß mehrstufige Auswahl- verfahren entstehen. Klumpenstichproben weichen am stärksten von einer rei- nen Zufallsauswahl ab und sind nur sehr selten repräsentativ. Es ist daher im Einzelfall genau zu prüfen, ob der ausgewählte Klumpen nicht die Fragestel- lung betreffende Verzerrungen aufweist, die seine Eignung als repräsentative Stichprobe in Frage stellen.

2.5 Mehrstufige Auswahl

Die bislang erläuterten Verfahren der geschichteten Stichprobe bzw. der Klum- penauswahl hatten gemeinsam, daß die Grundgesamtheit zunächst in Teil- gesamtheiten aufgeteilt wurde, aus denen nach verschiedenen Regeln die Un- tersuchungseinheiten in die endgültige Stichprobe ausgewählt wurden. Dieses Vorgehen läßt sich verallgemeinern. Für Fälle, in denen der Untersuchungs- bereich sehr groß ist und die Untersuchungseinheiten über z.B. ein sehr großes Territorium verteilt sind, stellt ein mehrstufiges Zufallsauswahlverfahren eine praktikablere Alternative zu den bislang vorgestellten Auswahlverfahren dar. Bei mehrstufigen Auswahlverfahren wird die Auswahl der eigentlich interes- sierenden Untersuchungs- d.h. Analyseeinheiten erst auf der letzten Stufe mehr- rerer, jeweils nachgelagerter Auswahlvorgänge vorgenommen. Analog zur Klumpenstichprobe unterscheidet man daher Auswahlseinheiten der 1. Stufe (Primäreinheiten), der 2. Stufe (Sekundäreinheiten) etc. In Zwischenschritten werden aus der Grundgesamtheit Teilmassen ausgewählt, aus diesen Teilmas- sen werden weitere Teilmassen gezogen usw., bis auf einer letzten Stufe dann die Beobachtungseinheiten aus den zuletzt gezogenen Teilmassen zufällig aus- gewählt werden. Mehrstufige Auswahlverfahren bestehen also aus einer Reihe nacheinander durchgeführter Zufallsstichproben, wobei die jeweils entstehende Zufallsstichprobe die Auswahlgrundlage der folgenden Zufallsstichprobe dar- stellt. Ein wesentlicher Aspekt der sukzessiven Erfassung von Teilmassen ist, daß eine Auswahlgrundlage immer nur für jede Stufe einzeln zu beschaffen ist, so daß man sich diesbezüglich auf praktisch zugängliche Listen mit Auswahl- einheiten auf jeder Stufe beschränken kann. Die Abbildung 2.7 verdeutlicht eine hierarchische Zerlegungsstruktur für den Fall einer dreistufigen Auswahl.

Die Grundgesamtheit wird wie bei der Klumpenstichprobe zunächst auf mehrere Urnen verteilt, von denen ein Teil zufällig ausgewählt wird. Im Ge-

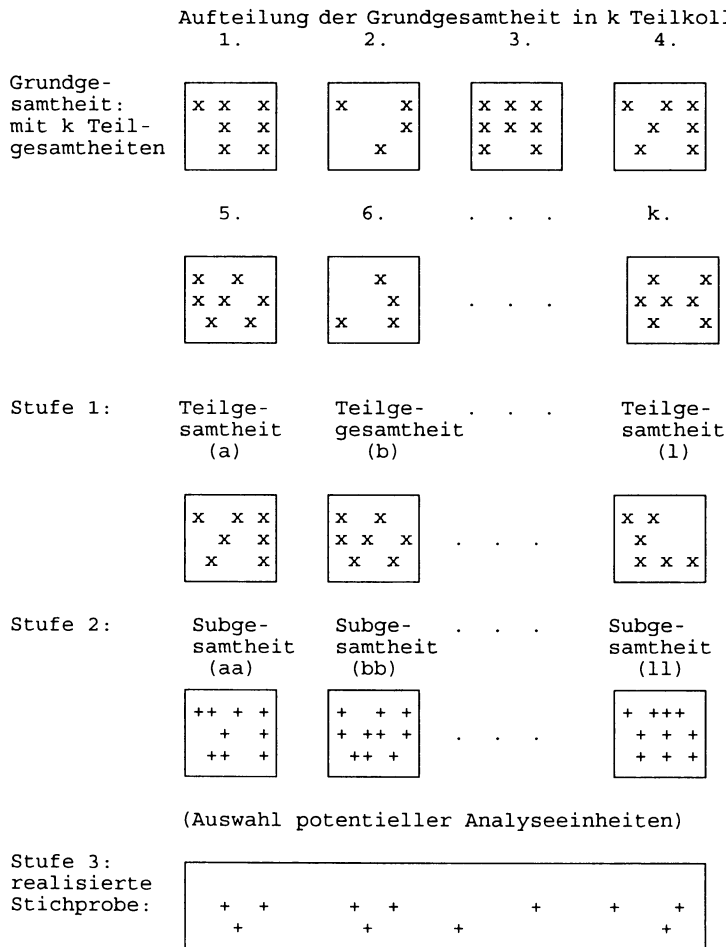


Abb. 2.7: Das Modell der mehrstufigen Auswahl

gensatz zur Klumpenauswahl bilden nun jedoch nicht alle Elemente der ausgewählten Urnen die Stichprobe, sondern nur diejenigen Elemente, die in einer zweiten Zufallsauswahl aus den ausgewählten Urnen entnommen worden sind.

Wie bei der Klumpenstichprobe wird auch die mehrstufige Auswahl hauptsächlich wegen der durch sie erzielbaren Aufwandsverringerung angewendet. Vor allem bei Verfahren mit mehreren Auswahlstufen ergibt sich gegenüber der einfachen Zufallsauswahl eine beträchtliche Reduktion der tatsächlich zu 'samplenden' Elemente (Abgrenzung, Auswahl nur für die Elemente der vorgela-

gerten Auswahlseinheiten). Auch wird erst auf der letzten Stufe des Auswahlverfahrens eine Liste der Untersuchungseinheiten erforderlich.

Ein Problem der mehrstufigen Auswahl ist die Bestimmung der optimalen Auswahlsätze (Festlegung der Größe der Auswahlseinheiten auf jeder Stufe). Erschwerend kommt hinzu, daß die Primäreinheiten in der Regel unterschiedlich viele Elemente enthalten. Für diesen Fall sind unterschiedliche Lösungsvarianten vorgeschlagen worden, die trotz ungleicher Größe der Primäreinheiten die gleichen Auswahlwahrscheinlichkeiten für jedes Element der Grundgesamtheit annähernd garantieren. Auf der ersten Stufe einer z.B. zweistufigen Auswahl werden Auswahlwahrscheinlichkeiten für die Primäreinheiten verwendet, die proportional zur Größe der Primäreinheiten sind. Auf der zweiten Stufe werden aus jeder ausgewählten Primäreinheit dieselbe Zahl von Sekundäreinheiten gezogen. Verfahren, die diesen Ansatz verwenden, werden »PPS-Designs« genannt (»Probability Proportional to Size«). Ihre Anwendung führt auch bei erschwerten Ausgangsbedingungen zu einer Zufallsauswahl aus der Grundgesamtheit.

Zusammenfassend läßt sich zu den zufallsgesteuerten Auswahlverfahren feststellen:

- Die Definition der Grundgesamtheit für eine empirische Untersuchung ist sehr problematisch und sollte sehr sorgfältig vorgenommen werden.
- Wird eine Stichprobe gezogen, um mit ihrer Hilfe Aussagen über die Grundgesamtheit zu treffen, muß die Stichprobe repräsentativ sein.
- Die Repräsentativität einer Stichprobe hängt von dem Stichprobenumfang und dem Auswahlverfahren ab. Klumpenstichproben sind am wenigsten repräsentativ. Reine Zufallsauswahlen, systematische Stichproben und geschichtete Stichproben liefern in der Praxis annähernd gleich »gute« Ergebnisse.

Bei der Darstellung der einzelnen zufallsgesteuerten Auswahlverfahren sollte deutlich werden, daß deren Anwendbarkeit von spezifischen Bedingungen abhängig gemacht werden muß und sich überdies jedes Auswahlverfahren durch bestimmte Vorteile einerseits und Nachteile andererseits charakterisieren läßt. Von einem in allen Fällen am besten geeigneten Auswahlverfahren kann daher nicht ausgegangen werden.

Die Beantwortung der Frage, unter welchen Bedingungen ein Auswahlverfahren der einen Klasse einem solchen der anderen Klasse vorzuziehen ist, hängt vom Zusammenwirken vieler Faktoren ab, letztlich jedoch auch vom Forschungsgegenstand, den Forschungsobjekten, dem Typ der Forschung, den Forschungszielen, den Forschungsressourcen und von den praktischen Gegebenheiten. Die Wahl des geeigneten Auswahlverfahrens kann daher immer nur im Hinblick auf die konkrete Aufgabenstellung getroffen werden.

Grundsätzlich besteht immer die Gefahr, daß durch das nicht konsequente Befolgen der Auswählerfordernisse Verzerrungen auftreten und sich damit Ein-

schränkungen hinsichtlich der Repräsentativität einer Stichprobe ergeben. Eine gute Möglichkeit zur wenigstens groben Überprüfung der Repräsentativität einer Stichprobe ist die Prüfung, ob die Stichprobe ähnliche Eigenschaften wie die Grundgesamtheit hinsichtlich zentraler Untersuchungsmerkmale aufweist. Voraussetzung dafür ist allerdings, daß entsprechende Daten über die Grundgesamtheit in Form von Häufigkeiten oder Anteilswerten in den untersuchungsrelevanten Merkmalsausprägungen zur Verfügung stehen.

2.6 Willkürliche und bewußte Auswahlen

Bei einer willkürlichen Auswahl (auch *Auswahl aufs Geratewohl* genannt) wird die Ziehung der Elemente aus der Grundgesamtheit nicht durch einen Auswahlplan kontrolliert. Dabei entscheidet gerade nicht ein kontrollierter Zufallsprozeß darüber, ob ein Element der Grundgesamtheit in die Stichprobe kommt oder nicht, sondern maßgebend ist ausschließlich die 'willkürliche' Entscheidung des Forschers, der die Auswahl vornimmt. Da weder die Grundgesamtheit sinnvoll definiert ist, noch vor der Stichprobenziehung jedes Element der Grundgesamtheit die gleiche (oder eine angebbare) Chance in Form einer Auswahlwahrscheinlichkeit hat, sind willkürliche Auswahlen für wissenschaftliche Zwecke nahezu immer wertlos. Hinzu tritt ihre fehlende intersubjektive Nachvollziehbarkeit.

Ein gutes Beispiel für die Probleme der willkürlichen Auswahl ist die Erhebung der amerikanischen Militärregierung über den Gesundheitszustand der deutschen Stadtbevölkerung nach dem 2. Weltkrieg. Bei dieser Untersuchung wurden in deutschen Großstädten an Straßenecken, Plätzen u.ä. Orten Personenwaagen aufgestellt. Vorbeikommende Passanten wurden gebeten, sich wiegen zu lassen. Auf diesem Wege sollte das Durchschnittsgewicht der deutschen Stadtbevölkerung ermittelt werden. Selbst wenn außer acht gelassen wird, ob das Gewicht ein in jeder Hinsicht überzeugender Indikator für den Gesundheitszustand ist oder nicht, lieferte die Untersuchung verzerrte Ergebnisse. Zum einen war die Wahrscheinlichkeit für mobile Personen in die Stichprobe zu gelangen höher als für kranke oder weniger mobile Personen. Zum anderen konnte an bestimmten Plätzen (z.B. Bahnhöfen) nicht ausgeschlossen werden, daß auch Personen aus ländlichen Gebieten gewogen wurden, die ihrerseits schwergewichtiger gewesen sein konnten.

Von der willkürlichen Auswahl unterscheidet sich die bewußte Auswahl (oder Auswahl nach Gutdünken) durch den Rückgriff auf vorhandene Informationen zur Grundgesamtheit allgemein und zu den Auswahl-elementen. Das Grundprinzip besteht darin, daß mit Sachkenntnis gezielt solche Elemente in die Stichprobe einbezogen werden, die für das Untersuchungsziel von großer Bedeutung sind, d.h. die die charakteristischen Merkmalsausprägungen der Elemente der Grundgesamtheit haben bzw. die für ein gegebenes Untersuchungsziel als typisch angesehen werden. Kellerer definiert die bewußte Auswahl unter Bezugnahme auf die Sachkenntnisse der Grundgesamtheit wie folgt:

»Bei der bewußten Auswahl entscheidet nicht der Zufall ... – im wahrscheinlichkeitstheoretischen Sinne – welche Elemente in die Teilerhebung einbezogen werden. ... Die Auswahl wird vielmehr auf Grund einer allgemeinen Kenntnis der Grundgesamtheit durch die mit der Planung und Durchführung der Erhebung betrauten Personen vorgenommen; ... auch bei einer bewußten Auswahl ist der Leitgedanke, sie so vorzunehmen, daß sie hinsichtlich der Untersuchungsmerkmale möglichst repräsentativ für die entsprechende Grundgesamtheit ist« (Kellerer, H.: 1960: Statistik im modernen Wirtschafts- und Sozialleben, Hamburg, S. 156). Die Zusammensetzung der Stichprobe wird somit unter Berücksichtigung von Informationen über die einzelnen Stichprobenelemente bewußt und gezielt gesteuert.

Ob ein Element der Grundgesamtheit ausgewählt wird, hängt vom Zutreffen vorher festgelegter (also angebbarer und intersubjektiv nachvollziehbarer) Kriterien ab. So muß z.B. die Grundgesamtheit angebbar sein, um überhaupt Kriterien für die kontrollierte Auswahl entwickeln zu können. Dabei kann sich die Kontrolle natürlich nur für einen Teil aller Erhebungsmerkmale erstrecken. Damit ein Element der Grundgesamtheit in die Stichprobe aufgenommen werden kann, muß es bestimmte Merkmale oder Merkmalskombinationen aufweisen. Entsprechend kann die bewußte, kontrollierte Auswahl auch immer nur für die zuvor festgelegten Merkmale oder Merkmalskombinationen repräsentativ sein (im Gegensatz zur Zufallsauswahl, bei der durch die Zufälligkeit der Entnahme automatisch eine repräsentative Auswahl im Hinblick auf alle Merkmale erreicht wird). Die Grenze zwischen bewußter Auswahl und Zufallsauswahl ist allerdings im Hinblick auf die Güte der Stichprobe fließend: Unter der idealtypischen Annahme, daß sowohl alle relevanten Merkmale zur Strukturierung der Grundgesamtheit als auch die Ausprägungen aller Elemente in diesen Merkmalen bekannt sind, läßt sich auch durch eine kontrollierte Entnahme der Stichprobenelemente eine der Zufallsauswahl äquivalente Stichprobe bilden. Die Festlegung der Merkmale richtet sich dabei nach den beabsichtigten Aussagen der Untersuchung; bewußte Auswahlen eignen sich am ehesten für Analysen mit eng eingegrenzten Fragestellungen. Die generelle Zielsetzung bei der Durchführung bewußter Auswahlen muß in einer möglichst weitgehenden Objektivierung des Verfahrens bestehen:

- exakte Abgrenzung von Grundgesamtheit und Auswahlinheit;
- Einengung des subjektiven Entscheidungsspielraumes durch die Verwendung umfassender empirischer Informationen zu Stichprobenelementen und Grundgesamtheit;
- Berücksichtigung der Untersuchungsfragen bei der Auswahl der Merkmale zur Strukturierung der Grundgesamtheit als Voraussetzung zur bewußten Auswahl.

Der Ausdruck »bewußte Auswahl« ist ein Oberbegriff für alle Auswahltechniken, deren Grundprinzip in einer hinreichenden Kenntnis der Grundgesamtheit besteht:

a) *Auswahl typischer Fälle*: Diese Auswahltechnik stellt die einfachste Variante von den bewußten Auswahlen dar, zugleich ist sie allerdings auch die problematischste. Es werden die Fälle ausgewählt, die als besonders »charakteristisch« (als besonders typisch) für die Grundgesamtheit angesehen werden. Maßgebend für die Kategorisierung ist die interne Homogenität der Elemente, die zu einem Typus gehören, sowie deren externe Isolierung von den anderen Elementen der Grundgesamtheit. Dabei besteht die Grundüberlegung darin, daß die Analyse auf diese relativ wenigen Elemente der Grundgesamtheit beschränkt werden soll. Wenn Untersuchungseinheiten hinsichtlich bestimmter zentraler Merkmale »typisch« sind für eine größere Gesamtheit von Untersuchungseinheiten, dann erlaubt deren Analyse auch Aussagen über die Grundgesamtheit insgesamt. Dabei ist zunächst festzulegen, hinsichtlich welcher Kriterien die Elemente »typisch« sein sollen, denn »typische« Elemente an sich gibt es nicht. Diese Kriterien können nur vom Untersuchungsziel – von den angestrebten Erkenntnisinteressen – her definiert werden. Dementsprechend werden auch nur Untersuchungseinheiten ausgewählt, die dem Untersuchungsziel entsprechen.

Eine Auswahl typischer Untersuchungseinheiten setzt entsprechende Vorkenntnisse über die Grundgesamtheit voraus. Man muß vor der Erhebung wissen, wie die relevanten Merkmale, nach denen die typischen Untersuchungseinheiten definiert werden, in der Grundgesamtheit verteilt sind. Um Aussagen über die Grundgesamtheit treffen zu können, muß dann bei der Auswahl typischer Fälle angenommen werden, daß die typischen Fälle in jeder Hinsicht den nicht ausgewählten Elementen gleichen. Diese Annahme ist für sozialwissenschaftliche Untersuchungen meist nicht realistisch und zudem nicht überprüfbar.

Meist sind die erforderlichen Vorkenntnisse über die Verteilung der (theoretisch) angemessenen Merkmale in der Grundgesamtheit nicht vorhanden oder nicht ohne weiteres feststellbar sein. Die Auswahl muß sich in diesem Fall an der Verwendung sog. Ersatzmerkmale orientieren. Über diese Ersatzmerkmale müssen ausreichende Kenntnisse vorhanden sein, sie muß zugleich leicht erfaßbar sein und man muß davon ausgehen, daß sie hoch mit den theoretisch interessierenden Merkmalen zusammenhängen (korrelieren). Nur dann ist die Annahme gerechtfertigt, daß die anhand von Ersatzmerkmalen ausgewählten typischen Fälle auch hinsichtlich der eigentlich interessierenden Merkmale typisch sind. Da diese Annahmen in der Regel nicht erfüllt sind, ist die Auswahl typischer Fälle in den meisten Fällen ein ungeeignetes Auswahlverfahren.

b) *Auswahl nach dem Konzentrationsprinzip*: Zu den bewußten Auswahlverfahren wird auch die Auswahl nach dem Konzentrationsprinzip (auch als *cut-off-Verfahren* bezeichnet) gezählt, das vor allem in der amtlichen Statistik häufig angewendet wird. Hierbei beschränkt man die Erhebung auf die für den Untersuchungsgegenstand besonders ins Gewicht fallenden Untersuchungseinheiten. Auch bei diesem Auswahlverfahren sind vorab Informationen über die

Auswahleinheiten notwendig, insbesondere über deren Bedeutsamkeit, gemessen an den jeweils im Vordergrund stehenden Untersuchungsmerkmalen. Besondere Bedeutung gewinnt dieses Auswahlverfahren in den Fällen, in denen ein relativ kleiner Teil der Grundgesamtheit einen großen Einfluß auf die untersuchten Merkmale ausübt. »Großer Einfluß« ist demnach das zentrale Kriterium, nach dem diejenigen Untersuchungseinheiten zu bestimmen sind, die in die Auswahl kommen sollen. »Bedeutsamkeit« kann sich auf die Wirkungen bestimmter Einheiten auf andere Einheiten beziehen (z.B. »die Machtelite« übt einen großen politischen Einfluß aus) oder ein großer Anteil an der Gesamtausprägung eines bedeutsamen Merkmals (z.B. Beschäftigtenzahl von Unternehmen). Ein Beispiel könnte darin bestehen, nur die (wenigen) Unternehmen mit dem größten Umsatz innerhalb eines bestimmten Wirtschaftssektors zu untersuchen.

Durch die Beschränkung auf relativ wenige, die Untersuchungsmerkmale besonders stark beeinflussende Untersuchungseinheiten läßt sich eine kleine, ökonomischere Stichprobe bilden, die trotzdem für einen großen Teil der Grundgesamtheit repräsentativ ist.

Die Auswahl nach dem Konzentrationsprinzip ist insbesondere bei solchen Fragestellungen angebracht, bei denen es weniger um die absoluten Werte von Untersuchungsmerkmalen geht (die 'absolute Niveaubestimmung'), als vielmehr um relative Unterschiede zwischen zwei oder mehreren Grundgesamtheiten (vergleichende Untersuchungen, Längsschnittuntersuchungen). Das Auswahlverfahren eignet sich nur bei sehr speziellen und eingeschränkten Fragestellungen.

c) Quoten-Auswahl: Eine spezielle Form der bewußten Auswahl wird vor allem in der Markt- und Meinungsforschung verwendet: das »Quota-Verfahren«. Die Auswahl nach Quoten strebt das gleiche Ziel an wie die Zufallsauswahl, d.h. die Auswahl eines repräsentativen Miniaturquerschnitts der Grundgesamtheit. Allerdings nimmt die Quoten-Auswahl im Vergleich zur Zufallsauswahl praktisch den gegensätzlichen Ausgangspunkt ein. Sie beginnt mit der Analyse und Aufstellung der statistischen Proportionen der Grundgesamtheit. Hierzu werden die Daten der amtlichen Statistik oder von bereits durchgeführten repräsentativen Erhebungen verwendet. Die Festlegung der Quotierungsmerkmale erfolgt anhand der bekannten prozentualen Verteilung relevanter Merkmale (Quotierungsmerkmale). Die zu erhebenden Merkmale in der Stichprobe müssen – empirisch oder zumindest theoretisch abgesichert – mit den Quotenmerkmalen hoch korrelieren. Die Untersuchungseinheiten werden derart in die Stichprobe aufgenommen, daß die Quotenmerkmale in der Stichprobe exakt in derselben Häufigkeit vorkommen wie in der Grundgesamtheit. Dadurch wird die Übereinstimmung der Struktur von Grundgesamtheit und Stichprobe als verkleinertes Abbild zumindest hinsichtlich der Quotenmerkmale erreicht. Da die Quotierung dem Verfahren der Schichtung bei der geschichteten Zufallsauswahl entspricht, kann man die Quoten-Stichprobe auch als eine »geschichtete bewußte Auswahl« bezeichnen.

Damit die Auswahl als eine maßstäbliche Verkleinerung der Grundgesamtheit bezeichnet werden kann, müssen die Proportionen allerdings hinsichtlich aller untersuchungsrelevanten Merkmale übereinstimmen. Da die hierzu notwendige Quotierung sehr aufwendig ist, versucht man nur jene Merkmale zur Quotierung zu verwenden, die einen hohen Zusammenhang (eine hohe Korrelation) mit den untersuchungsrelevanten Merkmalen aufweisen. Durch diese Korrelation – so die Annahme – ergibt sich auch für die nicht quotierten Merkmale eine Übereinstimmung mit den Proportionen in der Grundgesamtheit. Die Erfüllung dieser Forderung ist aber keinesfalls gesichert.

Bei der Festlegung der Quotierungsvorgaben beschränkt man sich in der Regel auf eine sehr begrenzte Zahl von Merkmalen, die leicht feststellbar sind (in der kommerziellen Umfrageforschung z.B. auf demographische Merkmale wie Alter, Geschlecht, Beruf oder Einkommen) und deren Verteilungen in der Grundgesamtheit aus der amtlichen Statistik hinlänglich bekannt sind.

Die Rechtfertigung der Anwendung des Quota-Verfahrens beruht meist auf praktischen Erwägungen, da dieses Verfahren weit kostengünstiger und zudem schneller durchzuführen ist als eine Zufallsauswahl. Das gilt insbesondere dann, wenn für die zu untersuchende Grundgesamtheit keine Liste vorliegt, aus der eine Zufallsstichprobe gezogen werden könnte. Die meisten Einwände lassen sich zusammengefaßt auf folgende Punkte konzentrieren:

- Das Quota-Verfahren stellt keine Zufallsauswahl dar; die Anwendung der Inferenzstatistik ist erschwert, wenn nicht unmöglich.
- Quotenmerkmale, die einen generellen Einfluß auf die relevanten Untersuchungsmerkmale ausüben und deren Verteilung in der Grundgesamtheit außerdem bekannt ist, lassen sich kaum identifizieren. Ein 'set' von Standardmerkmalen, das sich in der kommerziellen Umfrageforschung in Form von demographischen Merkmalen etabliert hat, existiert für wissenschaftliche Untersuchungen nicht.
- Aufgrund der Vielzahl der bei solchen Untersuchungen interessierenden Merkmale und aufgrund der Fülle unterschiedlicher Fragestellungen, werden sich Quotenmerkmale nur für konkrete Untersuchungen, abhängig von den jeweils zu untersuchenden Fragestellungen und Zielsetzungen, im Einzelfall benennen lassen. Selbst dann wird jedoch im Vergleich zur konventionellen Umfrageforschung eine höhere Zahl von Quotenmerkmalen notwendig sein, um eine annähernde Übereinstimmung der Struktur von Grundgesamtheit und Stichprobe zu sichern.

Zusammenfassung: Zur Beurteilung der durch die vorgestellten Auswahlverfahren zu erreichenden Repräsentativität der Stichprobe bietet die Gegenüberstellung in der Tabelle 2.2 eine erste Orientierung.

Der entscheidende Vorzug, den die zufallsgesteuerten Auswahlverfahren gegenüber bewußten Auswahlen aufweisen, besteht in der Erzeugung von repräsentativen Teilmassen. Sie stellen ein zwar verkleinertes, aber wirklichkeits-

Tab. 2.2: Zur Repräsentativität von Auswahlverfahren

Auswahlverfahren	Repräsentativität der Auswahl angestrebt durch:	Repräsentativität der Auswahl gesichert:
willkürliche Auswahlen	---	nein
bewußte Auswahlen	Informationen zu Stichprobenelementen und Grundgesamtheit	mit Einschränkungen oder nur für die direkt oder indirekt kontrollierten Merkmale
Zufallsauswahlen	zufällige Entnahme der Stichprobenelemente	ja

getreues Abbild der Grundgesamtheit dar. Daraus lassen sich unmittelbar folgende Vorzüge von repräsentativen Teilmassen gegenüber bewußt ausgewählten Teilmassen ableiten:

- Zufallsstichprobenergebnisse gestatten zuverlässige Schlüsse vom Teil aufs Ganze, sogenannte Repräsentationsschlüsse. Die statistische Schlußweise behandelt diese Problematik mit Hilfe des Begriffs der Wahrscheinlichkeit, mit dem jedes Stichprobenergebnis in Beziehung zu der entsprechenden Grundgesamtheit gesetzt wird. »Zuverlässig« ist dabei so zu verstehen, daß die Größe des Werteintervalls angebbar ist, in dem ein bestimmter, aus der Zufallsstichprobe ermittelter statistischer Kennwert (z.B. das arithmetische Mittel), in der Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit) liegen wird.
- Zufallsstichproben sind genauer, objektiver, indem sie die durch subjektive Ermessensentscheidungen hervorgerufene Gefahr systematischer Verzerrungen bei der Auswahl der Stichprobenelemente verhindern.
- Der Einsatz zufallsgesteuerter Auswahlverfahren ist die Voraussetzung für die Anwendung mathematisch-statistischer Testverfahren.
- Das Vorliegen von Zufallsstichproben ist eine notwendige Bedingung für die Vergleichbarkeit von Ergebnissen.

Wenn man zu diesen Vorteilen die Vorzüge von Teilerhebungen gegenüber einer Totalerhebung dazurechnet (weniger Aufwand und erheblich geringerer zeitlicher und finanzieller Ressourcenverbrauch), werden die mit den Zufallsstichprobenverfahren vorhandenen Möglichkeiten offensichtlich.

Dennoch dürfen einige Probleme nicht übersehen werden. Die Anwendung zufallsgesteuerter Auswahlverfahren darf nicht zum entscheidenden Kriterium für die methodologische Beurteilung einer Forschung gemacht werden, denn der Forschungserfolg wird maßgeblich von der theoretischen Durchdringung des ganzen Forschungsprozesses bestimmt. Mit der durch Stichprobenverfahren erreichten Populationsrepräsentanz und wesentlichen Aspekten der Merkmalsrepräsentanz sind nicht alle Probleme der Datenrepräsentanz gelöst.

Zufallsstichprobenverfahren lassen sich nicht auf sämtliche sozialwissenschaftliche Forschungsprobleme anwenden oder erweisen sich in manchen Fällen als unverhältnismäßig aufwendig. Dennoch braucht in derartigen Fällen nicht generell auf die Beantwortung einer Forschungsfrage verzichtet zu werden. Die aus nichtzufallsgesteuerten Stichproben erzielten Ergebnisse gelten allerdings nur zur Beschreibung der Stichprobe.

Ferner sollte auch nicht mit dem bloßen Nennen des Begriffs »Repräsentativität« automatisch der Anschein wissenschaftlicher Exaktheit verbunden werden. Letztlich muß davor gewarnt werden, die Repräsentativität einer Untersuchung mit ihrer Reichweite zu identifizieren. Während die Repräsentativität primär das Problem Stichprobe-Grundgesamtheit berührt, betrifft die Frage der Reichweite einer Untersuchung primär das Problem der Berechtigung der theoretischen Verallgemeinerung der Ergebnisse. Von den existierenden Verfahren der Stichprobenbildung werden in der Abbildung 2.8 die Grundtypen noch einmal zusammenfassend systematisiert.

3. Datensammlung und -aufbereitung für statistische Analysen

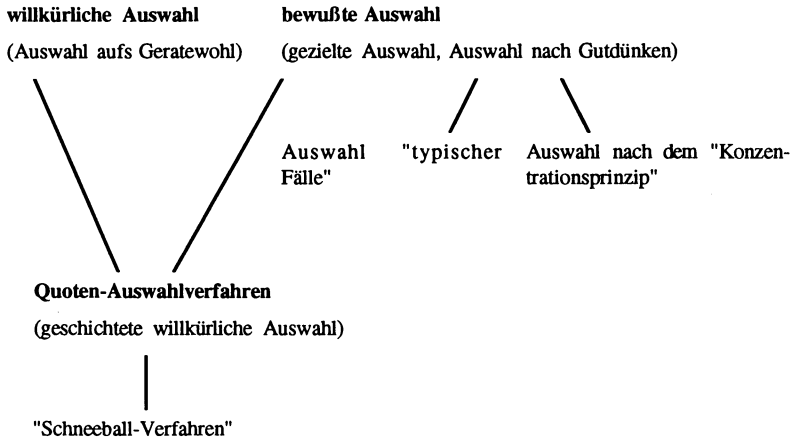
Um wissenschaftliche Aussagen unter Verwendung von statistischen Modellen überprüfen zu können, benötigt man *Daten* über den empirischen Gegenstandsbereich. Die auf der Grundlage begrifflicher Strukturierungen gemachten Beobachtungen werden zu Daten über ein Untersuchungsobjekt erst dadurch, daß sie in einer spezifischen, standardisierten Form registriert und aufbereitet werden.

3.1 Formaler Aufbau einer Datenmatrix

Daten sind die in geeigneter Form festgehaltenen und abrufbaren symbolischen Repräsentationen der an den Untersuchungseinheiten beobachteten Merkmale. Sie besitzen folgende formale Struktur (Galtung, J., 1967: *Theory and methods of social research*, London, S. 9ff):

- Die erhobenen Daten beziehen sich auf Untersuchungseinheiten. Die Untersuchungseinheiten sind diejenigen »Objekte«, auf die sich die untersuchungsleitenden oder zu prüfenden Aussagen/Hypothesen richten. Die Untersuchungseinheiten werden durchnummeriert: Mit n wird der Stichprobenumfang bezeichnet, der Index i drückt die Nummerierung der Untersuchungseinheiten aus (d.h. der Merkmalsträger $i = 1, \dots, n$).
- Die Daten beschreiben die Untersuchungseinheiten nicht in ihrer gesamten Komplexität, in ihrer »Ganzheit«, sondern lediglich im Hinblick auf ausgewählte Merkmalsdimensionen, bzw. genauer: den Merkmalen (kurz mit X symbolisiert). Die Anzahl der Merkmale sei mit K angegeben. Sämtliche Merkmale lassen sich dann allgemein mit dem Laufindex k ($k = 1, \dots, K$)

Übersicht: nicht zufallsgesteuerte Auswahlverfahren



Übersicht: zufallsgesteuerte Auswahlverfahren

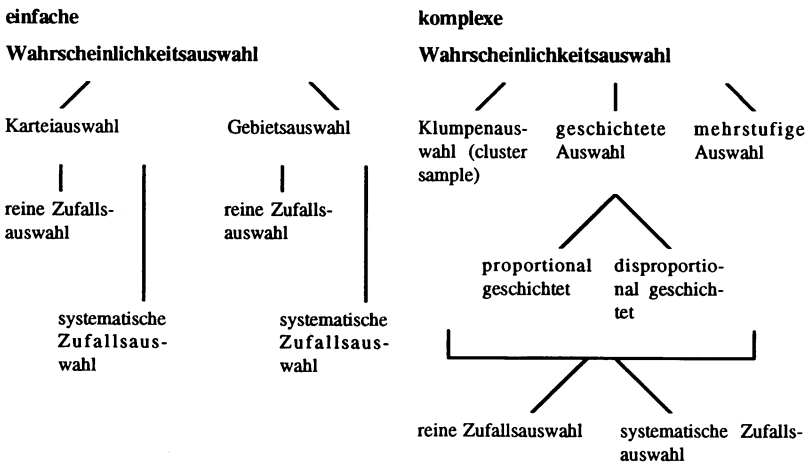


Abb. 2.8: Zusammenfassung der Typen von Auswahlverfahren

- wie folgt ausdrücken: $X_1, X_2, \dots, X_k, \dots, X_{K-1}, X_K$. Der Index j des k -ten Merkmals kann die Werte 1 bis m (Zahl der möglichen Ausprägungen des k -ten Merkmals) annehmen, wobei die Indices der Unterscheidung dienen.
- Beobachtet werden auf den interessierenden Merkmalsdimensionen die jeweiligen Ausprägungen für die Untersuchungseinheiten, d.h. die Merkmalswerte. Die Ausprägung des Merkmals X_k bei der Untersuchungseinheit i wird mit dem Kleinbuchstaben x_{ik} bezeichnet.

Tab. 3.1a: Formale Struktur einer Datenmatrix

Untersuchungs- einheiten (UE):	Variablenbündel:						
	X_1	X_2	X_3	\dots	X_k	\dots	X_K
UE ₁	x_{11}	x_{12}	x_{13}	\dots	x_{1k}	\dots	x_{1K}
UE ₂	x_{21}	x_{22}	x_{23}	\dots	x_{2k}	\dots	x_{2K}
.
.
.
UE _i	x_{i1}	x_{i2}	x_{i3}	\dots	x_{ik}	\dots	x_{iK}
.
.
.
UE _n	x_{n1}	x_{n2}	x_{n3}	\dots	x_{nk}	\dots	x_{nK}

Durch die spezifischen Ausprägungen je Untersuchungseinheit auf allen festgelegten Merkmalen wird jede Untersuchungseinheit *datenmäßig* charakterisiert. Um aus ihnen Informationen zur Klärung der interessierenden Fragen zu gewinnen, müssen die Beobachtungsdaten zunächst in ein bestimmtes Ordnungsschema gebracht werden. Dieser Aufbereitungsprozeß wird heute unter Nutzung der modernen Personal Computer-Technologie vorgenommen. Dazu müssen die Rohdaten zunächst in einem für die Datenverarbeitung geeigneten Format erfaßt werden. Die Tabelle 3.1a zeigt ein geeignetes Schema für eine allgemeine Datenmatrix in tabellarischer Form. Die empirischen Werte in der Datenmatrix werden auch als *Urliste* bezeichnet. Für jede Untersuchungseinheit i ($i = 1, \dots, n$) werden die erfaßten Meßwerte der k -ten Variablen ($k = 1, \dots, K$) spaltenweise in einer Zeile notiert, wobei die Reihenfolge festgelegt ist. Die Spalten, in die der Wert einer Variablen eingetragen wird, bezeichnet man als Feld. Die beobachteten Variablenausprägungen x_{ik} werden durch Eintragungen in diese Felder festgehalten. Die in diesem Format geordneten und auf einem Datenträger abgespeicherten Meßwerte einer Untersuchung bilden in ihrer Gesamtheit eine *Datendatei*.

Die Tabelle 3.1b zeigt beispielhaft eine überschaubare Datenmatrix aus einem bestimmten Forschungsfeld der Historischen Wahlforschung mit ausge-

wählten Sozialdaten auf der Untersuchungsebene der Wahlkreise in der Weimarer Republik, waagerecht die Merkmale, senkrecht die Untersuchungseinheiten (aus: Falter, J./Lindenberger, Th./Schumann, S. 1986: Wahlen und Abstimmungen in der Weimarer Republik. Materialien zum Wahlverhalten 1919–1933, München, S. 66). Diese Daten spiegeln lediglich einen Merkmalsausschnitt wider. Wir werden im folgenden den Gang der Darstellung beispielhaft auch anhand der Sozialstrukturmerkmale dieser Datenmatrix illustrieren, ergänzt durch die Wahlergebnisse der NSDAP bei den Reichtagswahlen 1930 bis 1933 (s. die Dokumentation der Wahlergebnisse in Falter/Lindenberger/Schumann a.a.O., S. 71–75).

Exkurs

Historische Wahlforschung: Fragestellung, Untersuchungseinheiten, Merkmale und Reichweite der Ergebnisinterpretation

Die Historische Wahlforschung untersucht allgemein die sozialen Bestimmungsfaktoren für die Entscheidungen bei politischen Wahlen etwa im Kaiserreich 1871 bis 1918 und in der Weimarer Republik 1919 bis 1933 unter Berücksichtigung der sozialstrukturellen Randbedingungen dieser Zeit. Ziel der Historischen Wahlforschung ist es, die Entwicklung der Parteien bzw. Parteigruppierungen bei Wahlen auf global nationaler, regionaler oder lokaler Ebene oder regional vergleichend zu beschreiben. Die Historische Wahlforschung beschäftigt sich darüberhinaus auch mit der zeitlichen Entwicklung von Wahlergebnissen; Beispiele sind der Vergleich von Wahlergebnissen über die Zeit oder die Analyse von Wählerwanderungen zwischen zeitlich benachbarten Wahlen. Neben der Frage, wie Parteien in interessierenden Gemeinden, Kreisen, Wahlkreisen, Großregionen bei konkreten Wahlen abschneiden, interessiert auch die unterschiedliche soziale, wirtschaftliche oder religiöse Struktur der genannten Gebietseinheiten. Eine von der Historischen Wahlforschung aufgeworfene Frage lautet z.B., wie erfolgreich die Parteien bei einer Wahl in Gebieten mit unterschiedlichen Prozentanteilen an Katholiken, Arbeitslosen, Arbeitern oder Selbständigen waren. Damit sind bestimmte gesellschaftliche Teilgruppen angesprochen, deren Stabilität und Wandel im Wahlverhalten untersucht werden können, wenn die zeitliche Entwicklung anhand aufeinanderfolgender Wahlen berücksichtigt wird. Ein Ziel der Historischen Wahlforschung ist u.a. für die Weimarer Republik die Beantwortung folgender Fragen:

- Woher kamen die Wähler der NSDAP?
- Wie groß war der Anteil der Arbeiterschaft an der Gesamtwählerschaft der NSDAP?
- Sind insbesondere jüngere Wähler gegenüber der NSDAP anfällig gewesen?

Tab. 3.1b: Beispiel einer Datenmatrix: Daten zur Sozialstruktur der Wahlkreise in der Weimarer Republik aus den Volkszählungen 1925 und 1933

Wkr. Nr.	Religion		Urbani- sierung		Sozial- struktur		Wirtschafts- struktur			Sozialstruktur in den Wirt- schaftsabtei- lungen				Krisen- indika- toren	
	rk	ev	l	s	ar	se	I	II	III ff.	ar/ I	ar/ II	se/ IIIff.	alo	Verschul- dung in I	
1. Ostpreußen	16	83	63	19	35	33	53	20	27	52	63	20	11	59	
2. Berlin	11	69	0	100	48	11	0	54	45	76	68	22	35	34	
3. Potsdam II	10	74	8	89	34	15	5	44	51	60	58	24	23	34	
4. Potsdam I	7	83	32	52	42	20	24	45	31	51	70	22	20	33	
5. Frankfurt O.	12	86	61	8	38	33	45	32	22	39	71	23	13	42	
6. Pommern	3	95	57	14	38	31	48	24	28	52	65	22	12	53	
7. Breslau	37	59	47	32	41	23	32	38	30	48	70	23	21	39	
8. Liegnitz	16	81	63	14	40	31	38	39	23	35	73	25	16	43	
9. Oppeln	89	10	53	27	39	29	40	37	23	30	76	20	20	28	
10. Magdeburg	5	89	45	26	42	24	32	39	29	53	72	25	17	27	
11. Merseburg	4	92	59	14	45	24	31	45	24	49	76	25	18	25	
12. Thüringen	7	86	57	12	44	25	28	49	23	30	72	30	19	19	
13. Schl.-Holstein	3	92	44	32	37	27	29	34	37	42	67	25	18	35	
14. Weser-Ems	26	71	40	30	34	37	38	28	33	24	66	25	13	20	
15. Osthannover	4	92	67	16	36	38	46	31	23	26	69	24	12	26	
16. S-Hann-Braun.	10	86	48	35	37	28	29	41	30	39	69	26	17	24	
17. Westfalen N.	53	45	37	35	44	26	26	50	24	20	76	26	19	17	
18. Westfalen S.	43	53	19	55	48	17	11	61	28	24	80	23	27	20	
19. Hessen-Nass.	28	68	54	35	34	33	31	38	31	15	70	24	18	18	
20. Köln-Aachen	83	15	28	44	40	23	18	47	35	31	72	25	22	14	
21. Koblenz-Trier	76	23	75	11	28	49	50	27	23	9	65	25	13	9	
22. Düsseldorf O.	44	49	2	85	47	13	3	63	34	33	73	25	28	25	
23. Düsseldorf W.	67	30	18	57	48	18	13	58	29	32	75	25	25	20	
24. O-Bayern-Schw.	88	10	52	34	33	34	37	33	30	29	64	28	12	27	
25. Niederbayern	96	4	80	6	29	48	60	23	17	20	64	32	10	28	
26. Franken	48	51	59	25	33	39	40	39	22	16	66	32	13	20	
27. Pfalz	42	56	57	17	40	33	33	46	21	21	71	32	19	9	
28. Dresden-Btz.	5	87	39	34	44	19	16	55	29	45	73	28	24	33	
29. Leipzig	3	83	30	52	45	17	12	55	35	58	72	27	25	24	
30. Chemnitz-Zw.	3	90	33	29	52	17	8	71	21	34	71	33	24	32	
31. Württemberg	33	66	58	19	34	39	41	39	20	8	65	30	9	15	
32. Baden	58	39	55	29	35	36	35	40	25	10	67	25	16	11	
33. Hessen-Dst.	31	65	53	26	37	33	33	41	26	16	70	26	18	13	
34. Hamburg	5	78	3	93	39	14	3	34	63	62	67	22	29	27	
35. Mecklenburg	4	94	52	29	38	27	39	26	34	60	64	25	13	39	
Im Deutschen Reich	32	63	44	36	39	27	29	42	29	34	69	26	19	27	

Merkmalsbeschreibung: »wkr«: Amtliche Nummer des Wahlkreises; »rk«: Prozentsatz der römisch-katholischen Bevölkerung; »ev«: Prozentsatz der evangelischen Bevölkerung; »l«: Prozentsatz der Bevölkerung in Gemeinden mit bis zu 5.000 Einwohnern; »s«: Prozentsatz der Bevölkerung mit mehr als 50.000 Einwohnern; »ar«: Prozentsatz der Arbeiter und erwerbslosen Arbeiter an allen Berufspositionen; »se«: Prozentsatz der Selbständigen und Mithelfenden an allen Berufsgruppen; »I«: Prozentsatz der in der Landwirtschaft Erwerbstätigen an allen Erwerbstätigen; »II«: Prozentsatz der in der Industrie und Handwerk Erwerbstätigen an allen Erwerbstätigen; »IIIff«: Prozentsatz der in den übrigen Wirtschaftssektoren Erwerbstätigen an allen Erwerbstätigen; »ar/I«: Prozentsatz der Arbeiter und ihrer Angehörigen an den Berufszugehörigkeiten in der Landwirtschaft; »ar/II«: Prozentsatz der Arbeiter und ihrer Angehörigen an den Berufszugehörigkeiten in Industrie und Handwerk; »ar/IIIff.«: Prozentsatz der Selbständigen und Mithelfenden sowie ihrer Angehörigen an den Berufszugehörigkeiten der übrigen Wirtschaftsabteilungen; »alo«: Prozentsatz der Arbeitslosen an allen Erwerbspersonen; »Verschuldung in I«: Schulden der landwirtschaftlichen Betriebe in Prozent des Einheitswerts.

- Verließ die regionale Ausbreitung der NSDAP für das gesamte Reich homogen oder breitete sich die NSDAP regional bei Wahlen unterschiedlich aus?
- Hat die hohe Arbeitslosigkeit die NSDAP-Wahlerfolge wesentlich mitbestimmt?
- Gibt es Unterschiede zwischen den konfessionellen Trägergruppen hinsichtlich der NSDAP-Wählerschaft?
- Wie groß ist der Einfluß der Verschuldung in Landwirtschaft und Gewerbe auf die NSDAP-Wahlerfolge?

Die Beantwortung dieser Fragen setzt die Messung und Erhebung der relevanten Merkmale voraus. Das Studium von Wählerverhalten kann auf zwei unterschiedlichen Untersuchungsebenen erfolgen.

1. Individualebene: Die Erhebungsmerkmale werden auf der Ebene des einzelnen Wählers erhoben. Die Zuordnung von individuellen Wahlentscheidungen und erklärenden Merkmalen – z.B. soziodemographische Angaben, politische Grundüberzeugungen, persönliche Wertvorstellungen, Beurteilung politischer Programme, Einstellungen zu Parteien, etc. – erfolgt auf der Ebene des Wahlberechtigten. Die erhobenen Daten geben Auskunft über Merkmalsausprägungen von Individuen in einer Stichprobe aus den Wahlberechtigten einer Grundgesamtheit. Die moderne empirische Wahlforschung gewinnt diese Daten mittels repräsentativer Bevölkerungsumfragen.

2. Aggregatebene: Dem historischen Wahlforscher stehen keine Informationen auf der Ebene des einzelnen Wahlberechtigten zur Verfügung. »Die für das Kaiserreich und die Weimarer Republik vorliegenden Informationen sind lediglich in Form sogenannter Aggregatdaten zugänglich, d.h. als gebietsmäßig aufbereitete offizielle Statistiken. Informationen über das Wahlverhalten sind daher beispielsweise nur in Form von Wahlergebnissen verfügbar, die auf der Ebene bestimmter Gebietskörperschaften – z.B. Gemeinden, Land- und Stadtkreise oder höhere Verwaltungsgliederungen – ausgewiesen sind. Das gleiche gilt für die das Wahlverhalten 'erklärenden' Merkmale, also Volkszählungsdaten, Wirtschaftsstatistiken etc. Sie werden vom historischen Wahlforscher mit den Wahldaten in Beziehung gesetzt, um beispielsweise Informationen über das Wahlverhalten der verschiedenen sozialen Gruppen zu erhalten« (Falter, J., 1991: Hitlers Wähler, München, S. 54). Da die Daten der Sozial-, Wirtschafts- und Wahlstatistik nach räumlichen Kriterien aggregiert sind, bezeichnet man sie auch als ökologische Daten. Zwar beruhen diese Daten ursprünglich auf Individualzählungen, man kann jedoch nicht mehr auf einzelne Individuen zurückgreifen. Für die Weimarer Republik sind im Rahmen des großen Forschungsprojektes »Wählerbewegungen zum Nationalsozialismus« unter der Leitung von Jürgen Falter sowohl Wahl- als auch Wirtschafts- und Sozialdaten auf unterschiedlichen Aggregationsniveaus dokumentiert. Die Datenmatrix in Tab. 3.1b illustriert ein hohes räumliches Aggregationsniveau: Merkmalsaus-

prägungen in Prozentanteilen auf Wahlkreisebene (N = 35). Auf einem wesentlich niedrigerem Aggregationsniveau bewegt sich demgegenüber der Kreis- und Gemeindedatensatz.²

Als strukturelle Faktoren, die für die Wahlentscheidungen, die langfristige Wählerbindung wie die Wählerwanderung von besonderer Bedeutung sind, gelten in der Historischen Wahlforschung insbesondere die konfessionellen, die sozioökonomischen und die sozialgeographischen Verhältnisse räumlich abgegrenzter Einheiten wie z.B.:

- Prozentanteil der evangelischen bzw. römisch-katholischen Bevölkerung,
- Prozentanteil der in den Wirtschaftssektoren Beschäftigten sowie das Verhältnis zwischen Selbständigen und abhängig Beschäftigten,
- Anteil der Gemeindegrößenklassen.

Die Wahlergebnisse – etwa die NSDAP-Stimmenanteile bei den Reichstagswahlen 1930 bis 1933 – werden dann gemeinsam mit den Wirtschafts- und Sozialdaten in geeignete statistische Analysen einbezogen, um etwa den Zusammenhang zwischen Berufsstruktur und NSDAP-Stimmenanteile festzustellen oder die Wirkung alternativer Erklärungsfaktoren – wie z.B. die Konfessionsverteilung oder die Gemeindegröße – auf die Entwicklung der NSDAP-Stimmenanteile zu bestimmen.

Da in wahlhistorischen Studien von Aggregatdaten ausgegangen wird, ist die Frage zu klären, welches Aggregationsniveau der Datenanalyse zugrunde gelegt werden sollte. Geht man von den 35 Wahlkreisen der Weimarer Republik als territoriale Einheiten aus, sind weitreichende Aussagen aus differenzierten statistischen Analysetechniken aufgrund der geringen Zahl der Untersuchungseinheiten nicht ableitbar. Studien auf der Grundlage hochaggregierter Daten verdecken die strukturellen Disparitäten innerhalb dieser territorialen Einheiten. Als analytische Einheiten sind die 35 Wahlkreise intern zu heterogen, um weiterreichende Untersuchungen etwa über den Zusammenhang zwischen Sozialstrukturmerkmalen und Stimmenanteilen einzelner Parteien zu erlauben. Aus diesem Grunde bewegen sich die Auswertungen in dem Forschungsprojekt »Wählerbewegungen zum Nationalsozialismus« ausschließlich auf der Ebene der Kreis- und Gemeindeeinheiten des Reiches.³

² Einen Eindruck von der komplexen Datenstruktur der Wahl-, Wirtschafts- und Sozialdaten, die die Verknüpfung von Aggregatdaten unterschiedlicher Bezugsgrößen (Wahlkreise, Gemeinden etc.) nahelegte, läßt sich der entsprechenden Dokumentation von Dirk Hänisch entnehmen (Hänisch, D., 1989: Inhalt und Struktur der Datenbank »Wahl- und Sozialdaten der Kreise und Gemeinden des Deutschen Reiches von 1920 bis 1933«, in: Historical Social Research, 14, S. 39–67). Für weitergehende (Sekundär-) Analysen sind der wissenschaftlichen Öffentlichkeit die Wahl- und Sozialdaten auf der Ebene der Stadt- und Landkreise über das Zentralarchiv für Empirische Sozialforschung, Abteilung Zentrum für Historische Sozialforschung, zugänglich (ZA-Nr. 8013: Wahl- und Sozialdaten der Kreise und Gemeinden des Deutschen Reiches von 1920 bis 1933).

³ In dem Buch »Hitlers Wähler« präsentiert Jürgen Falter eine umfassende Gesamt-

Für die in diesem Abschnitt darzustellenden, elementaren statistischen Modelle ist diese Datenbasis der 35 Wahlkreise durchaus geeignet, den mit statistischen Methoden nicht vertrauten Historiker anhand eines überschaubaren Datensatzes die Versteh- und Nachvollziehbarkeit des Übergangs von den Daten einer Datenmatrix zu statistischen Modellen zu erleichtern. Die Beschreibung der Wahlergebnisse auf Wahlkreisebene erlaubt zumindest eine nähere regionale Eingrenzung des Wahlgeschehens innerhalb der für den Wahlakt eigens geschaffenen administrativen Einheiten. In dem letzten Abschnitt werden wir vergleichend zwei unterschiedliche Aggregationsniveaus berücksichtigen.

Abschließen wollen wir diesen Exkurs mit einem Hinweis auf die Reichweite der Aussagen, die im Rahmen von Aggregatdatenanalysen getroffen werden können. Generell ist bei der Interpretation statistischer Ergebnisse die Wahl der Erhebungseinheit und damit die Untersuchungsebene zu berücksichtigen. Da in der Historischen Wahlforschung die erforderlichen Daten nicht auf der Ebene des einzelnen Wählers erhoben werden können, stehen lediglich Daten über die Stimmen und die Prozentanteile der einzelnen Parteien und sozialstatistische Merkmale in bestimmten, geographisch definierten Erhebungseinheiten zur Verfügung. Die interessierenden Fragen bewegen sich allerdings auf der Individualdatenebene, z.B. die Beantwortung der Frage, ob die Arbeitslosen in der Weimarer Republik tendenziell eher die NSDAP gewählt haben. Die Historische Wahlforschung kann zunächst die Frage beantworten, ob in Gemeinden mit einem hohen Arbeitslosenanteil tendenziell ein hoher Stimmenanteil für die NSDAP einhergeht. Diese Aussage ist jedoch nicht identisch mit dem Abstimmungsverhalten von Personen, die in einem bestimmten Wahljahr arbeitslos waren und ihre Stimme zugunsten oder zuungunsten der NSDAP abgegeben haben. In der Historischen Wahlforschung fallen die angestrebte Aussageebene (individuelles Wählerverhalten) und die Untersuchungsebene (Wahlverhalten in Gemeinden) auseinander. Aus der Mißachtung dieses Sachverhalts können bei einem Rückschluß von Ergebnissen auf der Aggregatebene auf die Individualebene erhebliche Fehler entstehen. Angesprochen ist die Problematik des *ökologischen Fehlschlusses*. »Der Fehler der naiven Disaggregation besteht darin, derartige Zusammenhänge zwischen politischen und sozialstatistischen Merkmalen, die auf der Ebene von Gebietseinheiten festgestellt werden, so zu interpretieren, als spiegelten sie notwendigerweise und ohne Verzerrung Zusammenhänge auf der Individualebene wider. Dies darf aber . . . nicht als selbstverständlich vorausgesetzt werden« (Falter, a.a.O., S. 56). In zahlreichen Untersuchungen zur Problematik des ökologischen Fehlschlusses wurde nachgewiesen, daß die auf Gebietsebene beobachteten Zusammenhänge zwischen Merkmalen im allgemeinen sehr viel größer sind als die entsprechenden individuellen Zusammenhänge. Weiter hat sich erwiesen, daß Aggregat-

darstellung der Einzelergebnisse zu den nationalsozialistischen Wahlerfolgen am Ende der Weimarer Republik, die auf der Ebene der Gemeinden oder Kreise und kreisfreien Städte des Reiches gewonnen wurden (Falter, a.a.O.).

und Individualdatenergebnisse umso stärker auseinanderfallen, je größer die verwendeten Gebietseinheiten gewählt werden. Die Verwendung von möglichst kleinen Gebietseinheiten (d.h. Aggregationsstufen) eröffnet zumindest Möglichkeiten für genauere Schätzungen von Zusammenhängen auf der Individualenebene⁴.

Als analytische Untersuchungseinheiten (UE's) sind in diesem Abschnitt beispielhaft die 35 Wahlkreise der Weimarer Republik ($i = 1, \dots, 35$; Zeilen der Datenmatrix) durch die wichtigsten Strukturmerkmale X_k ($k = 1, \dots, 14$; Spalten der Datenmatrix) nach dem Stande der Volkszählung (VZ) von 1925 bzw. 1933 charakterisiert. Es handelt sich um eine Totalerhebung mit einer eindeutigen räumlichen und zeitlichen Abgrenzung. Bezieht man die Wahlkreis-Nummerierung als Merkmal in die Datenaufnahme mit ein, ergibt sich eine 35×15 -Datenmatrix, die in Form der Tabelle 3.1b auch auf einer PC-Festplatte gespeichert werden kann.⁵ Diese Form der Datenmatrix wird von den meisten statistischen Softwarepaketen zur Abspeicherung unterstützt.

Die einzelnen Merkmale sind in unserem Beispiel auf einer einheitlichen Skala erhoben worden: Angaben in Prozent, d.h. die Merkmalswerte der Datenmatrix stellen Prozentanteile dar. Der Eintrag x_{12} (= 16) bedeutet z.B. die Ausprägung des zweiten Merkmals (Prozentsatz der römisch-katholischen Bevölkerung), die an der ersten Untersuchungseinheit (= Ostpreußen) erhoben wurde. Die 10-te Zeile in der Datenmatrix gibt alle für *Magdeburg* erhobenen Merkmalswerte an, und die 5-te Spalte der Datenmatrix enthält die Ausprägungen des Merkmals *Prozentsatz der Bevölkerung in Städten mit mehr als 50.000 Einwohnern* für sämtliche Untersuchungseinheiten.

Die Datenorganisation in einer Datenmatrix unterliegt grundsätzlich den für das Messen schon genannten Zuordnungskriterien (Kromrey, H., 1991: Empirische Sozialforschung, 5., überarb. u. erw. A., S. 167):

- *Vollständigkeit*: Es dürfen in der Datenmatrix keine Felder x_{ik} leer bleiben. In den Fällen, in denen auf einem Merkmal ein Wert empirisch nicht ermittelt werden kann, ist dieses Ergebnis durch die Formulierung einer zusätzlichen Merkmalsausprägung »fehlender Wert« (»missing value«) kenntlich zu machen. Diese Merkmalswerte sind keine Meßwerte und werden daher bei der Datenauswertung nicht berücksichtigt.

⁴ Zusammenfassend s. Pappi, F.U., 1977: Aggregatdatenanalyse, in: Koolwijk, J. van/Wieken-Mayser, M. (Hrsg.), 1977: Techniken der empirischen Sozialforschung, Band 7, München, S. 78–110).

⁵ Das Zentralarchiv für Empirische Sozialforschung, Abteilung Zentrum für Historische Sozialforschung, verwendet im Rahmen der EDV-Ausbildung in den Herbstseminaren den Gesamtdatensatz auf Wahlkreisebene (d.h. die Sozialstrukturmerkmale einschließlich sämtlicher Wahlergebnisse von der Wahl zur Nationalversammlung 1919 bis zur Reichstagswahl 1933). Zum eigenen Nachvollzug der im folgenden dargestellten Ergebnisse stellen wir auf Wunsch diesen Übungsdatensatz auf einer Diskette zur Verfügung.

- *Vergleichbarkeit*: Die Untersuchungsbedingungen müssen für alle Untersuchungseinheiten UE_i ($i = 1, \dots, n$) gleich sein. Dies besagt, daß die Merkmalsdimensionen X_k in der Datenmatrix auch tatsächlich Eigenschaftsdimensionen der Untersuchungseinheiten sind. Vergleichbarkeit besagt ferner, daß die Untersuchungseinheiten immer nur hinsichtlich eines einzigen Merkmals miteinander verglichen werden können.
- *Klassifizierbarkeit*: Für jedes Paar (UE_i, X_k) muß genau ein Wert x_{ik} existieren, d.h. jeder Untersuchungseinheit UE_i ($i=1, \dots, n$) muß mindestens eine Merkmalsausprägung zugewiesen werden. Zugleich darf keiner Untersuchungseinheit mehr als eine Ausprägung auf jedem Merkmal zugeordnet werden.

Die Transformation von Daten aus Erhebungsbögen oder schriftlichen Dokumenten in statistisch verarbeitbare Daten nennt man Verschlüsseln (oder Ver-coden): Um Informationen auf einen Datenträger übertragen zu können, ist es notwendig, eindeutige Regeln für die Beziehung zwischen einer Information und ihrer Darstellung auf einem Datenträger anzugeben. Diese Verschlüsselung der Daten erfolgt nach einem zuvor von den Forschern entwickelten Verschlüsselungsplan (Codeplan). Der Codeplan gibt an, welche Zeichen/Ziffern den Ausprägungen der Merkmale zugeordnet werden sollen. Er enthält ferner die Angabe der Spaltenposition für die Ausprägungen der einzelnen Variablen.

Das Verschlüsseln und die Erfassung der Daten für die PC-Aufbereitung erfordert je nach Art der Erhebung unterschiedlich viele Schritte. In den meisten Fällen wird zunächst eine Verschlüsselung auf dem Erhebungsbogen selbst erfolgen. Für diesen Zweck wird auf dem Erhebungsbogen eine Codeleiste am Rand vorgesehen. Die Daten aus der Codeleiste überträgt man dann auf die Festplatte eines PC. Die mehrfache Übertragung der Daten ist sehr arbeitsaufwendig und fehlerträchtig. Übertragungsfehler können bei jedem Schritt auftreten. Deshalb wird in der Praxis versucht, die Zahl der Schritte möglichst zu reduzieren und Sicherungen gegen Fehler einzubauen (z.B. durch doppelte Datenerfassung und späteren Vergleich). Bei einfacher Datenlage ist es möglich, die Daten direkt von der Codeleiste des Erhebungsbogens oder sogar unmittelbar aus dem Erhebungsbogen auf das Speichermedium zu übertragen. Sind schwierige Verschlüsselungsvorgänge zu bewältigen, ist dieses verkürzte Verfahren allerdings sehr fehleranfällig.

Zur Erleichterung der Erfassung kann es in einfachen Fällen möglich sein, die Verschlüsselung gleich bei der Erhebung vorzunehmen. Dann tragen die erhebenden Personen die empirischen Beobachtungswerte sofort in eine Schlüsselleiste ein. In den letzten Jahren haben sich zunehmend computergestützte Erfassungs- und Verschlüsselungstechniken durchgesetzt. Der erste Schritt besteht darin, daß die bereits verschlüsselten Daten mit Hilfe eines speziellen Computerprogramms auf das Speichermedium übertragen werden. Das Datenerfassungsprogramm *SPSS DATA ENTRY* ist ein Beispiel dafür; alternativ können auch Datenbankprogramme wie z.B. dBASE eingesetzt wer-

den. Der Vorteil dieses Verfahrens liegt darin, daß ein großer Teil möglicher Übertragungsfehler durch entsprechende Kontrollen bei der Eingabe vermieden wird. Man kann z.B. den Bereich gültiger Werte für die einzelnen Variablen angeben; Eingaben außerhalb dieses Bereichs werden von dem Programm nicht akzeptiert.

3.2 Tabellarische Darstellung univariater Häufigkeitsverteilungen

In der Datenaufbereitungsphase einer statistischen Untersuchung und für die Darstellung der statistischen Informationen werden geeignete Darstellungstechniken benötigt, die dem Betrachter die Ergebnisse möglichst übersichtlich präsentieren sollen. Zunächst werden Methoden für die Behandlung eines einzigen Merkmals (*univariate* Betrachtungsweise) vorgestellt. Obwohl in aller Regel bei einer quantitativen Untersuchung mehrere Merkmale erhoben werden, sind in dem ersten Auswertungsschritt für jedes einzelne Merkmal, d.h. für jede einzelne Spalte der Datenmatrix

x_{11}	x_{12}	x_{13}	\dots	x_{1k}	\dots	x_{1K}
x_{21}	x_{22}	x_{23}	\dots	x_{2k}	\dots	x_{2K}
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
x_{i1}	x_{i2}	x_{i3}	\dots	x_{ik}	\dots	x_{iK}
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
x_{n1}	x_{n2}	x_{n3}	\dots	x_{nk}	\dots	x_{nK}

spezifische Darstellungstechniken anzuwenden. Univariate Auswertungen liefern schon wesentliche Aufschlüsse über die Struktur der Daten und weisen gegebenenfalls auf Unstimmigkeiten in den erhobenen Daten hin.

Der Übergang von der Urliste zur tabellarischen Darstellung statistischer Informationen markiert die erste Phase statistischer Informationsverdichtung. Eine *Tabelle*, d.h. eine *geordnete, komprimierte Darstellung der Rohdaten*, ist die grundlegende statistische Wiedergabeform. Bei der Aufbereitung von Daten ist man zunächst daran interessiert, wie häufig die einzelnen Ausprägungen eines Merkmals X_k in der Stichprobe vom Umfang n auftreten. Beobachtet man an n Untersuchungseinheiten ein Merkmal X_k , das in m Ausprägungen vorkommt, so sind die *absoluten Häufigkeiten* wie folgt zu notieren:

Häufigkeiten $h_1, h_2, \dots, h_j, \dots, h_m$ der m Ausprägungen eines Merkmals X_k .

Bei der Aufbereitung von Daten in Häufigkeitstabellen sind bei metrischen Merkmalen in der Regel Gruppen (Klassen) von Ausprägungen zu bilden. Da-

bei ist jedoch zu beachten, daß mit jeder Klassenbildung ein gewisser Informationsverlust einhergeht. Häufigkeitstabellen fallen in der Datenaufbereitungsphase zunächst in Form von Druckprotokollen an, die mit Hilfe eines Statistik-Programmpakets erzeugt werden. Diese *Arbeitstabellen* sind in der Wiedergabe des statistischen Materials in der Regel unzureichend gestaltet, als das sie direkt für eine Veröffentlichung verwendet werden können. *Veröffentlichungstabellen* müssen gewisse Standards erfüllen, damit sie aus sich heraus lesbar sind. Die Abbildung 3.1 illustriert ein Schema für die tabellarische Darstellung von Häufigkeiten.

Überschrift 1)

Vorspalten

Lfd. Nr.	Spaltentext	Spaltenübergreifender Text			Summen- spalte
		Spaltentext	Spaltentext	...	
1	2	3	4
1	Zeilentext				
2	Zeilentext				
3	Zeilentext		Tabellen-		
...	...		felder		
Summen- zeile					

Tabellenkopf

1) Fußnote

Abb. 3.1: Allgemeines Schema für eine Tabelle
 Quelle: Hippmann, H.-D., 1994: Statistik für Wirtschafts- und Sozialwissenschaftler, Stuttgart, S. 47.

Da für die verschiedensten Zwecke Tabellen aufgestellt werden, ist es nicht möglich, starre Regeln für die Tabellenform anzugeben. Es können jedoch einige Grundelemente benannt werden, die eine tabellarische Darstellung enthalten muß, damit sie auch für jeden Betrachter lesbar ist. Als Bausteine zur Erstellung einer Tabelle können folgende allgemeine Regeln angegeben werden:

- 1) Jede Tabelle muß mit einer Überschrift versehen sein, die über den sachlichen Inhalt, den Zeitraum oder Zeitpunkt der Erfassung sowie den örtlichen Geltungsbereich informiert.
- 2) Tabellen müssen eine Kopf- und eine Vorspalte enthalten. Diese kennzeichnen die genauen Inhalte der einzelnen Spalten und Zeilen, d.h. den Merkmalsnamen und die Merkmalsausprägungen.
- 3) Eine Tabelle kann eine Summenzeile enthalten. Diese muß auch als solche kenntlich gemacht werden. Hierbei sollten verständliche Begriffe wie »Summe« oder »Insgesamt« verwendet werden.
- 4) Durch die Unterteilung in den Tabellenspalten und -zeilen entstehen die einzelnen Tabellenfelder. Umrahmt von Vorspalte und Tabellenkopf liegt

der Zahlenteil der Tabelle, der die statistischen Informationen enthält. Die Tabellenfelder des Zahlenteils dürfen bei fehlenden Informationen nicht leer bleiben, sondern müssen durch »-« oder andere Sonderzeichen besetzt und gegebenenfalls erläutert werden.

- 5) Zur Ergänzung und Erläuterung der Tabelle, einzelner in der Tabelle verwendeten Bezeichnungen und Abkürzungen oder für Anmerkungen zu einzelnen Tabellenfeldern können Fußnoten verwendet werden. Fußnoten schreibt man direkt an das Ende der Tabelle (und nicht an das Ende der Seite, auf der die Tabelle steht!).
- 6) Sollen in Fußnoten oder in einem kommentierenden Textteil einzelne Zeilen oder Spalten sowie Tabellenfelder direkt angesprochen werden, so bietet sich zur besseren Orientierung in der Tabelle eine Numerierung der einzelnen Tabellenspalten und -zeilen an.
- 7) An das Ende der Tabelle gehört ferner eine Quellenangabe. Ist z.B. die Tabelle das Resultat eigener Zählungen oder Berechnungen, so kann man angeben: eigene Erhebung oder eigene Berechnung. Grundsätzlich muß die Herkunft aller Zahlen einer Tabelle dokumentiert sein.

Die inhaltliche Tiefengliederung einer Tabelle kann unterschiedliche Informationen umfassen. Eine besonders wichtige Art der Gliederung folgt aus der Beschreibung einer univariaten Häufigkeitsverteilung. Die linke Spalte einer Häufigkeitstabelle enthält dementsprechend die Anordnung der Merkmalsausprägungen x_j ($j = 1, \dots, m$). Bei einem ordinalen Merkmal wird die Größenstruktur als Ordnungskriterium für die Bezeichnung der Tabellenzeilen herangezogen. Liegt ein nominales Merkmal vor, können allein sachliche Kriterien eine Reihenfolge der Ausprägungen festlegen. Zu jeder möglichen Merkmalsausprägung ist die Häufigkeit des Auftretens anzugeben. In der zweiten Spalte steht daher zu jedem Merkmalswert x_j der Variablen X_k die absolute Häufigkeit h_j (oft mit »f« abgekürzt; engl.: frequency; in Formelausdrücken auch als n_j notiert). Die Verteilung der Häufigkeiten auf alle unterschiedlichen Ausprägungen x_j ($j = 1, \dots, m$) der Variablen X_k nennt man die *absolute Häufigkeitsverteilung*. Die Summe der zweiten Spalte ($h_1 + h_2 + h_3 + \dots + h_m$) entspricht genau der Anzahl der Untersuchungseinheiten n in der Stichprobe. Für Häufigkeitsvergleiche vorteilhafter ist die Angabe von *relativen Häufigkeiten*. Sie werden aus h_j/n berechnet. Prozentzahlen gewinnt man, indem der Ausdruck h_j/n mit 100 multipliziert wird. Diese normierte und auf 100 bezogene relative Häufigkeit drückt somit für jede Ausprägung x_j ($j = 1, \dots, m$) aus, wieviel Prozent der Untersuchungseinheiten in der Stichprobe genau die Merkmalsausprägung x_j besitzen. Sie werden in die vierte Spalte eingetragen; ihre Summe muß dann 100% ergeben. Die Tabelle 3.2a illustriert allgemein den formalen Aufbau einer Häufigkeitsverteilung für ein k -tes Merkmal X_k aus der Datenmatrix, die Tabelle 3.2b illustriert die Ergebnisse der Wahl zur Nationalversammlung 1919 für den Wahlkreis Bremen. Untersuchungseinheiten sind die abgegebenen gültigen Stimmen.

Tab. 3.2a: Häufigkeitsverteilung eines diskreten Merkmals

Merkmals- ausprägung x_j von X_k	absolute Häufigkeit h_j (n_j)	relative Häufigkeit h_j/n	Prozent $(h_j/n)*100$
x_1	h_1 (n_1)	h_1/n	$(h_1/n)*100$
x_2	h_2 (n_2)	h_2/n	$(h_2/n)*100$
x_3	h_3 (n_3)	h_3/n	$(h_3/n)*100$
.	.	.	.
.	.	.	.
.	.	.	.
x_j	h_j (n_j)	h_j/n	$(h_j/n)*100$
.	.	.	.
.	.	.	.
.	.	.	.
x_m	h_m (n_m)	h_m/n	$(h_m/n)*100$
Summen- zeile:	n	1.00	100%

Tab. 3.2b: Die Wahl zur Nationalversammlung 1919 im Wahlkreis Bremen:
Stimmanteile absolut und in Prozent der gültigen Stimmen

Partei	absolute Häufigkeit h_j	relative Häufigkeit h_j/n	Prozent $(h_j/n)*100$
KPD	11.358	0,077	7,7
USPD	28.565	0,193	19,3
SPD	48.576	0,327	32,7
DDP	29.483	0,199	19,9
DVP	19.896	0,134	13,4
WV	10.496	0,070	7,0
Summe:	148.374	1.00	100

Abkürzungen:

KPD: Kommunistische Partei Deutschlands, USPD: Unabhängige Sozialdemokratische Partei Deutschlands, SPD: Sozialdemokratische Partei Deutschlands, DDP: Deutsche Demokratische Partei, DVP: Deutsche Volkspartei, WV: Wirtschaftliche Vereinigung.

Lesart:

Von den 148.374 abgegebenen gültigen Stimmen entfielen 48.576 auf die SPD; korrespondierende Anteilswert von 0,327 entspricht einem Prozentanteil von 32,7%.

Quelle: Falter, J./Lindenberger, Th./Schuman, S., 1986: Wahlen und Abstimmungen in der Weimarer Republik, München, S. 93.

Tab. 3.3a: Häufigkeitsverteilung eines klassifizierten Merkmals

Größenklasse K_j von dem Merkmal X_k	absolute Häufigkeit h_j	relative Häufigkeit h_j/n	relative Häufigkeit in Prozent $(h_j/n)*100$
$x_1^U - x_1^O$	h_1	h_1/n	$(h_1/n)*100$
$x_2^U - x_2^O$	h_2	h_2/n	$(h_2/n)*100$
$x_3^U - x_3^O$	h_3	h_3/n	$(h_3/n)*100$
.	.	.	.
.	.	.	.
.	.	.	.
$x_j^U - x_j^O$	h_j	h_j/n	$(h_j/n)*100$
.	.	.	.
.	.	.	.
.	.	.	.
$x_m^U - x_m^O$	h_m	h_m/n	$(h_m/n)*100$
Summen- zeile:	n	1.00	100%

Tab. 3.3b: Gemeindegrößenklassen in der Weimarer Republik

Größenklassen der Gemeinden in Tausend Einwohner	Gemeinden: absolute Häufigkeit h_j	Gemeinden: relative Häufigkeit h_j/n	Gemeinden: relative Häufigkeit in Prozent $(h_j/n)*100$
bis 2	792	0,197	19,7
2 - 5	2190	0,526	52,6
5 - 20	907	0,217	21,7
20 - 100	207	0,050	5,0
über 100	71	0,017	1,7
Summe:	4167	1.00	100

Quelle: Falter, J./Lindenberger, Th./Schuman, S., 1986: Wahlen und Abstimmungen in der Weimarer Republik, München, S. 174.

Bei einem ordinal-skalierten Merkmal können die einzelnen Merkmalsausprägungen in eine Reihenfolge gebracht werden. Das höhere Skalenniveau erlaubt die Tabellierung zusätzlicher Häufigkeitsdarstellungen. Oftmals interessiert, wieviele Untersuchungseinheiten unter einer bestimmten Ausprägungsgrenze liegen. Zu diesem Zweck werden zusätzlich *aufwärtskumulierte absolute und relative Häufigkeiten* angegeben. Die Summenhäufigkeit H_j gibt die Häufigkeit der Untersuchungseinheiten an, die die erste bis j -te Ausprägung aufweisen:

$$H_j = h_1 + h_2 + h_3 + \dots + h_j .$$

Die aufwärtskumulierten relativen Häufigkeiten erhält man nach Division durch die Gesamtzahl der Merkmalsträger:

$$H_j / n, \quad j = 1, \dots, m .$$

Die Häufigkeitsverteilung wird dann unübersichtlich, wenn ein Merkmal in sehr vielen Ausprägungen auftritt. Die tabellarische Darstellung würde nicht mehr hinreichend die Funktion der Informationsverdichtung erfüllen. In solchen Fällen betrachtet man nicht mehr die einzelnen Ausprägungen, sondern faßt verschiedene Ausprägungen in *Klassen* zusammen. Dies ist insbesondere bei metrisch-skalierten Merkmalen erforderlich. Mit der Zusammenfassung von Merkmalsausprägungen entstehen klassifizierte Häufigkeitsverteilungen mit den Klassen K_j ($j = 1, \dots, m$). Die Klassenuntergrenze der j -ten Klasse sei mit x_j^U bezeichnet, die Klassenobergrenze mit x_j^O . In der Tabelle 3.3a ist eine allgemeine Häufigkeitstabelle mit einer Klasseneinteilung dargestellt. Die Tabelle 3.3b illustriert ein Beispiel für klassifizierte Daten. Untersuchungseinheiten sind die Gemeinden der Weimarer Republik, das auszuwertende Merkmal deren Einwohnerzahl zum Zeitpunkt 1928. Die Gemeinden sind in Gemeindegrößenklassen zusammengefaßt.

Die Klasseneinteilung spielt eine wichtige Rolle bei der Darstellung der statistischen Informationen. Die Zusammenfassung von Einzelinformationen in Klassen führt zu einer höheren Übersichtlichkeit. Andererseits bewirkt eine Klassifizierung den Verlust der Einzelinformationen. Eine ungeeignete Klasseneinteilung und damit eine ungeeignete Zusammenfassung von Einzelinformationen kann zu falschen Aussagen und Schlüssen führen. Zu klären sind folgende Fragen: Wieviele Klassen sind sinnvollerweise zu bilden? Welche Klassenbreite ist geeignet? Für die Klassenbreite wird üblicherweise bei allen Klassen der gleiche Wert verwendet (konstante Klassenbreite). Vielfach legt jedoch der empirische Befund, d.h. die ursprüngliche Verteilung der Merkmalsausprägungen oder inhaltliche Gesichtspunkte unterschiedliche Klassenbreiten nahe. Die Tabelle 3.3b ist ein Beispiel für ungleiche Klassenbreiten. Mit zunehmender Einwohnerzahl wird die Zahl der Gemeinden immer niedriger (ab 20.000 Einwohner sind es lediglich 278 Gemeinden). Hier empfiehlt es sich, von größeren Klassenbreiten auszugehen.

3.3 Grundlegende graphische Darstellungsformen

Die *statistische Graphik* ist neben der Tabelle eine zweite wichtige Form der Darstellung statistischer Informationen. Graphische Darstellungen können statistische Tabellen begleiten oder für sich allein anstelle einer Tabelle stehen. Die Graphik ist die kürzere, einprägsamere und übersichtlichere Form der Wiedergabe von Tabellenwerten, aber sie geht in der Regel auf Kosten von Detailinformationen. Graphische Darstellungen haben allerdings die Eigenschaft, die Leistungsfähigkeit der menschlichen Informationsverarbeitung besser auszunutzen als irgendein anderes Medium der Informationsvermittlung. Besonders hervorzuhebende Ergebnisse sollten daher neben einer ausführlicheren tabellarischen Darstellung noch einmal in einer Graphik konzentriert werden.

Das Gebiet der statistischen Graphik hat sich inzwischen zu einem breiten Anwendungsfeld entwickelt. Eine strukturelle Einteilung statistischer Graphiken ist notwendig, um einen Überblick über die Vielzahl graphischer Darstellungsformen zu vermitteln. Neben reinen Präsentationsgraphiken für die Ergebnisdarstellung gibt es Graphiken mit eher datenanalytischen Charakter (*Datenanalysegraphiken*). Schließlich gibt es noch das Gebiet der *graphischen Methoden* als eigenständiges statistisches Forschungsinstrumentarium. Wir beschränken die folgende Darstellung jedoch nur auf Präsentationsgraphiken.

Der Konstruktionsaufwand war bis vor einigen Jahren noch ein gewichtiger Nachteil. Der Einsatz leistungsfähiger PC's in Verbindung mit flexibler Graphik-Software hat diesen Nachteil inzwischen völlig aufgehoben. Die Vielzahl statistischer Graphiken hat in Verbindung mit dem modernen PC-Einsatz bei der Erstellung von statistischen Diagrammen inzwischen eine weite Verbreitung gefunden. Mit Hilfe entsprechender Software ist es möglich, sehr schnell aufwendige Graphiken zu erzeugen.¹ Flexible Veränderungen in einer graphischen Darstellung ermöglichen individuelle Darstellungsformen. Darüberhinaus bietet die Verwendung der Graphik-Software auch die Möglichkeit, die dritte Dimension für die Ergebnisdarstellung zu nutzen.

Grundsätzlich gibt es drei Grundelemente zur graphischen Darstellung statistischer Informationen: *Punkte, Linien und Flächen*. Mit diesen 'Graphikbausteinen' lassen sich Punkt-, Linien- und Flächendiagramme entwickeln, die sich fünf Grundtypen zuordnen lassen: Säulen-, Kurven-, Punkte-, Kreis- und Balkendiagramme. In der Abbildung 3.2 sind diese fünf Graphikformen gegenübergestellt.

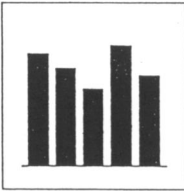
Die Werte einer Tabelle beinhalten immer eine Kernaussage, die in einen spezifischen Vergleich ausgedrückt werden kann. Der spezifische Vergleich ist dann in eine geeignete Graphik zu transformieren. Die Wahl des geeigneten

¹ Das Angebot statistischer Graphikpakete ist vielfältig. Das derzeit mächtigste Programm ist Harvard Graphics for Windows. Es folgen die Programme Charisma for Windows, Axum und Astound for Windows 1.5. Darüberhinaus verfügen zahlreiche Statistik-Programmpakete inzwischen über eine integrierte Graphik (z.B. SPSS für Windows, Statistica, Systat für Windows usw.).

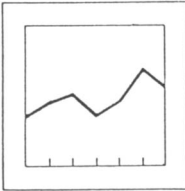
Tab. 3.4: Grundtypen von Vergleichen und Graphikformen

Grundtypen von Vergleichen:	Grundtypen von Diagrammformen
Struktur	Kreis
Rangfolge	Balken
Zeitreihe	Säule Kurven (Linien)
Häufigkeit	Säule
Korrelation	Punkt

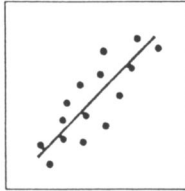
Säulen-
diagramm



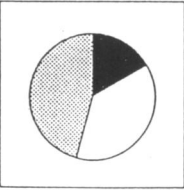
Kurven-
diagramm



Punkte-
diagramm



Kreis-
diagramm



Balken-
diagramm



Abb. 3.2: Grundtypen von Präsentationsgraphiken
Quelle: Zelazny, G., 1989: Wie aus Zahlen Bilder werden: Wirtschaftsdaten überzeugend präsentieren, Wiesbaden, S. 25.

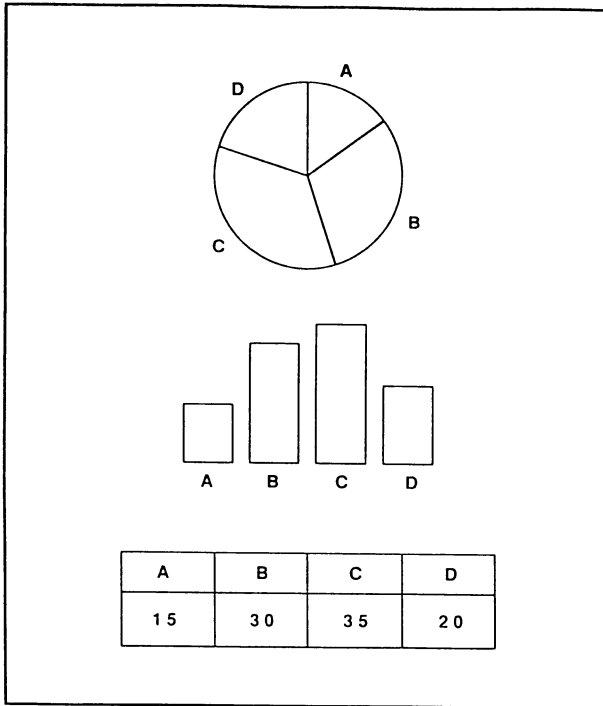


Abb. 3.3: Darstellung von Prozentanteilen als Flächenproportionen
Quelle: Fox, J./Long, J.S., 1990: Modern methods of data analysis, 2nd. edition, Newbury Park, S. 22.

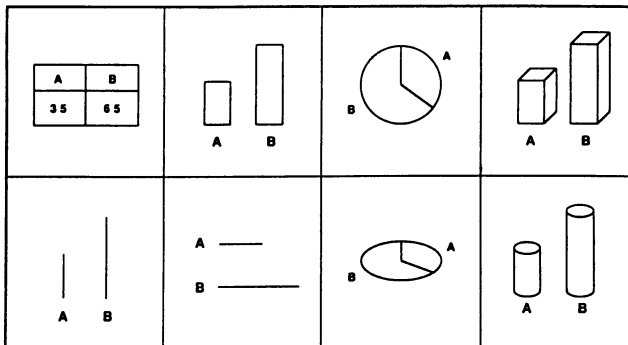


Abb. 3.4: Unterschiedliche graphische Möglichkeiten zur Veranschaulichung von Prozentanteilen
Quelle: Fox, J./Long, J.S., 1990: Modern methods of data analysis, 2nd. edition, Newbury Park, S. 24.

Graphik-Typs kann in drei Schritte angegeben werden (s. Zelazny, G., 1989: Wie aus Zahlen Bilder werden: Wirtschaftsdaten überzeugend präsentieren, Wiesbaden, S. 25):

- Von Tabellenwerten zur Aussage: Welche Aussage?
- Von der Aussage zum Vergleich: Welcher Vergleich?
- Vom Vergleich zum Diagramm: Welche Form?

Man unterscheidet fünf Grundtypen von Vergleichen: Struktur, Rangfolge, Zeitreihe, Häufigkeit und Korrelation. Sie korrespondieren direkt mit den Grundformen von Präsentationsgraphiken, d.h. die Vergleiche erfordern jeweils eine der fünf Graphikformen.² Die Zusammengehörigkeit von Aussage und Graphiktyp verdeutlicht die Tabelle 3.4.

Mit *Kreisdiagrammen* können Prozentanteile als Fläche von Kreissektoren dargestellt werden. Das graphische Darstellungsmittel ist daher die Fläche eines Kreisausschnitts, die den jeweiligen statistischen Wert repräsentiert (Proportionalität zwischen dem Winkel des Sektors und dem Prozentanteil, wobei die Winkelsumme 360° der Summe von 100% entspricht). Die Darstellung von Zahlen im Kreisdiagramm ist selbsterklärend und die Informationen auch für den Nichtstatistiker ohne zusätzliche Erklärungen leicht nachvollziehbar. Kreisdiagramme sind dann ein eher ungünstiges Darstellungsmittel für Prozentanteile, wenn die Anzahl der Merkmalsausprägungen größer Sechs ist, da die Winkel optisch viel schwieriger vergleichbar sind. Nur wenn die Sektoren mit den Prozentwerten bezeichnet werden, kann der Vergleich einfacher ausfallen. Dennoch sollte ein Kreisdiagramm nur gewählt werden, wenn ein Merkmal nicht mehr als sechs Ausprägungen aufweist.

Ein alternatives Flächendiagramm ist die *Stab- oder Säulengraphik*. Oberhalb der – auf einer Horizontalen angeordneten – Merkmalsausprägungen werden Säulen gezeichnet, deren Höhe sich proportional zu den darzustellenden Häufigkeiten bzw. Prozentanteilen verhalten. In der Abbildung 3.3 sind die beiden Flächendiagrammtypen anhand eines einfachen Zahlenbeispiels gegenübergestellt. Bei der Konstruktion eines Säulendiagramms ist darauf zu achten, daß sämtliche Säulen gleich breit sind und die Höhenskala bei Null beginnt. Die Säulen sollten nicht zu dicht nebeneinanderliegen.

Die Abbildung 3.4 vermittelt einen Überblick über Variationsmöglichkeiten von Flächendiagrammen. So können etwa die Stäbe in einem Stabdiagramm im Extremfall lediglich als Striche gezeichnet werden. Für den Betrachter ergeben sich dadurch große Unterschiede für die Dekodierung der statistischen Information. Darstellungen, die in dem unteren Abbildungsteil aufgenommen wurden, sollten eher vermieden werden.

² Säulen-, Balken- und Kreisdiagramm lassen sich auch dreidimensional darstellen. Wir haben auf diese – an sich informationsleere Information – in der folgenden Darstellung verzichtet

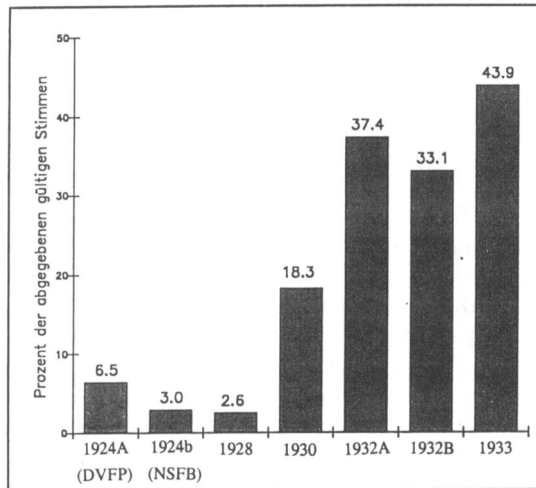


Abb. 3.5: Säulendiagramm für den Vergleich von zeitlich geordneten Beobachtungen
 Quelle: Falter, J., 1991: Hitlers Wähler, München, S. 21.

Das Säulendiagramm wird bevorzugt zur graphischen Darstellung von absoluten Häufigkeiten herangezogen. Darüberhinaus eignet sich dieser Diagrammtyp auch zur Wertedarstellung von kurzen Zeitreihen mit bis zu maximal zehn Zeitpunkten. Unter einer *Zeitreihe* versteht man eine zeitlich geordnete Folge von Merkmalsausprägungen. Auf der horizontalen X-Achse werden die Zeitpunkte abgetragen (Zeitachse, z.B. Jahresangaben, Jahresintervalle, Monatsdaten etc.). Jede Säule auf der Zeitachse repräsentiert dann einen zeitspezifischen Beobachtungswert eines Merkmals. Die Höhe der Säulen gibt jeweils proportional die numerische Größe der darzustellenden Merkmalswerte an. Als Beispiel ist in der Abbildung 3.5 die Entwicklung des NSDAP-Anteils bei den Reichstagswahlen 1924 bis 1933 graphisch dargestellt. Dieser Graphiktyp ist insbesondere dann für die Veranschaulichung zeitpunktspezifischer Informationen geeignet, wenn auf der Zeitachse die Zeitintervalle liegen, in denen das abzubildende Merkmal aus inhaltlichen Gründen nicht beobachtet werden kann (weil z.B. Wahlen lediglich zu bestimmten Zeitpunkten stattfinden). Säulendiagramme werden auch nur dann zur Darstellung von Zeitreihendaten verwendet, wenn der Verlauf (oder Trend) der Daten nicht im Vordergrund steht.

Eine weit verbreitete statistische Graphik zur Veranschaulichung von Zeitreihen ist das *Linien-* bzw. *Kurvendiagramm*, bei denen über die horizontale Zeitachse die Merkmalswerte einer Zeitreihe als Symbole (z.B. ».«, »+«, »*«, etc.) markiert und durch einen Linienzug verbunden werden, damit der Betrachter einen Eindruck von der zeitlichen Entwicklung eines Merkmals bekommt. Zur optischen Unterstützung ist es empfehlenswert, bei Kurvendia-

grammen ein sogenanntes Gitternetz zu verwenden, das aus horizontalen und vertikalen Referenzlinien besteht. Sie dienen dem Betrachter als zusätzliche graphische Stützen für ein einfaches Zuordnen der Punkte zu den an den Achsen des Koordinatensystems abgetragenen Skalenwerten (Merkmalswerte und Zeitpunkte). Diese Darstellungsform für Zeitreihendaten zeigt insbesondere bei sehr vielen Zeitpunkten am deutlichsten, ob ein bestimmter zeitlicher Trend vorliegt. Die Abbildung 3.6a veranschaulicht als Beispiel die nationale Verschuldung Englands von 1688 bis 1800 in Form eines Kurvendiagramms, die Abbildung 3.6b zeigt die natürliche Bevölkerungsbewegung in der Bundesrepublik Deutschland und in der DDR 1946 bis 1988.

Werden die einzelnen Merkmalsausprägungen entlang der horizontalen Zeitachse durch Geradenstücke verbunden, dann wird die zeitliche Entwicklung der Zeitreihe besonders hervorgehoben. Diese Vorgehensweise hat aber auch einen Nachteil. Sie geht von Annahmen über den Verlauf zwischen den Zeitabschnitten aus. So wurde in den Liniendiagrammen der Abbildung 3.6b ein linearer Verlauf zwischen den Zeitpunkten angenommen. Diese Annahme verliert an Bedeutung, je kürzer die Zeitintervalle erhoben werden. Im Fall von Wirtschaftsdaten, die in der Regel saisonale Schwankungen aufweisen, wird z.B. von Monatsdaten ausgegangen, da eine Liniengraphik mit Jahreswerten diese Schwankungen vernachlässigt.

Vielfach werden in einem Diagramm mehrere Zeitreihenmerkmale gleichzeitig dargestellt (*Mehrfach-Liniendiagramm*, s. Abbildung 3.6b). Neben den Achsenbeschriftungen ist in dieser Art von Graphik vor allem die deutliche Beschriftung der verschiedenen Linien wichtig. Als Alternative kann man auch unterschiedliche Linientypen verwenden (z.B. durchgezogene, gepunktete, gestrichelte Linien), deren inhaltliche Bedeutung in einer gesonderten Legende im Diagramm erläutert werden muß.

Bei einem Rangfolge-Vergleich ist ein *Balkendiagramm* zur graphischen Darstellung von Häufigkeiten oder Prozentanteilen der geeignete Graphiktyp. Diese Darstellungsform ist vom Prinzip her mit dem bereits beschriebenen Stab- oder Säulendiagramm identisch, nur daß die Säulen horizontal angeordnet sind. Das Balkendiagramm ist daher ein um 90° gedrehtes Säulendiagramm. Ein Balkendiagramm bietet sich für die Darstellung der nach Größe geordneten Rangfolge der einzelnen Häufigkeiten oder Prozentanteile einer Häufigkeitstabelle an. Das folgende Beispiel ist einen Ausschnitt aus den Ergebnissen der Wahlen zur Nationalversammlung am 19. Januar 1919 auf Wahlkreisebene. Graphisch veranschaulicht werden in der Abbildung 3.7 die SPD-Anteile ausgewählter Wahlkreise. Wie in der Abbildung zu sehen ist, wurden die Wahlkreise entsprechend ihren SPD-Anteilen geordnet. Die Größenverhältnisse sind optisch in einem geordneten Balkendiagramm weitaus besser zu erkennen als in einem Säulendiagramm. Der zweite Grund für die Verwendung eines Balkendiagramms ist rein praktischer Natur. Links neben den Balken steht hinreichend Raum für die Bezeichnungen der einzelnen Merkmalsausprägungen zur

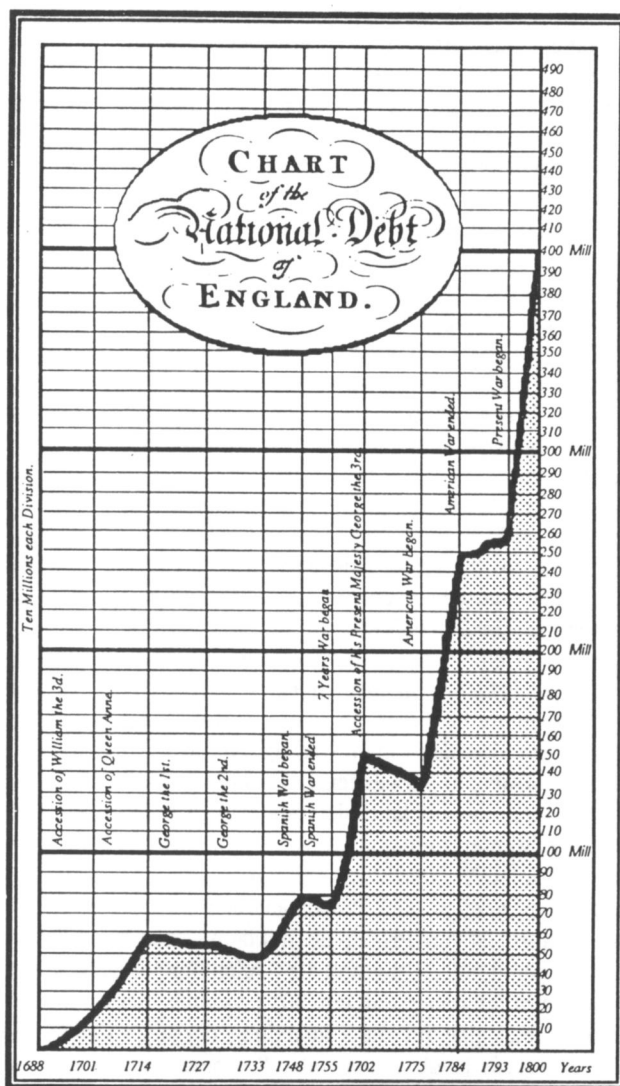


Abb. 3.6a: Kurvendiagramm: Die nationale Verschuldung in England von 1688 bis 1800

Quelle: Fox, J./Long, J.S., 1990: Modern methods of data analysis, 2nd. edition, Newbury Park, S. 15.

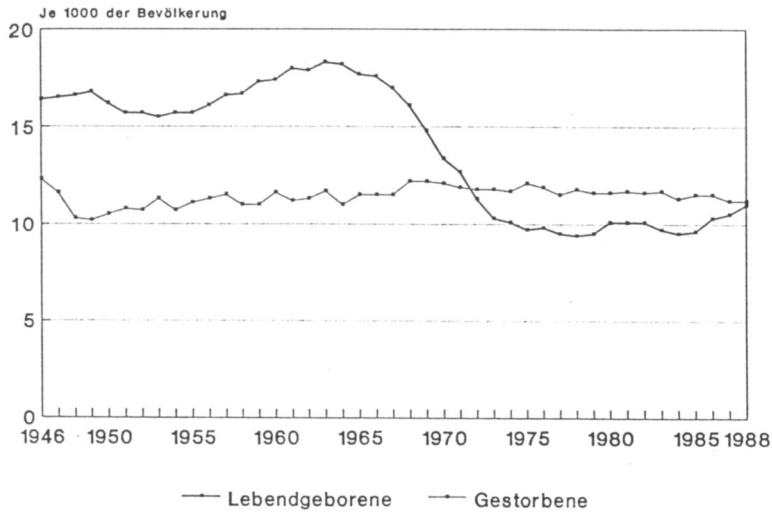


Abb. 3.6b: Natürliche Bevölkerungsbewegung in der Bundesrepublik Deutschland 1946 bis 1988 (je 1000 der Bevölkerung)

Quelle: Ritter, G.A./Niehuss M., 1991: Wahlen in Deutschland 1946-1991, München, S. 31.

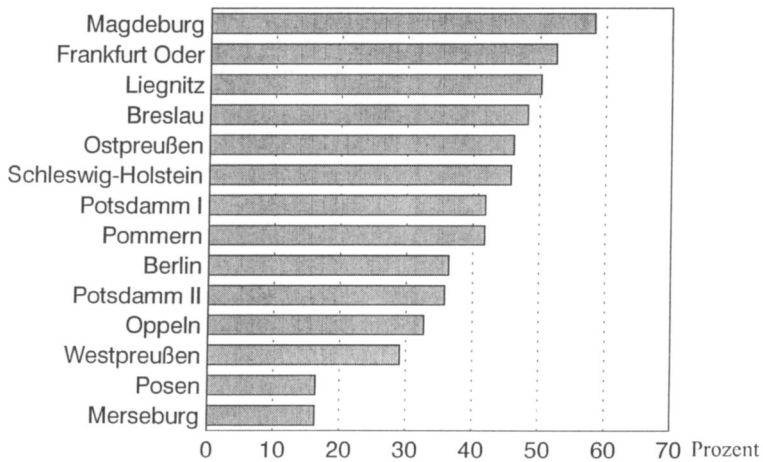


Abb. 3.7: Beispiel für einen Rangfolge-Vergleich

Quelle: Werte aus: Falter, J. u.a., 1986: Wahlen und Abstimmungen in der Weimarer Republik, München S. 67.

Verfügung, beim Säulendiagramm dagegen muß man sich gegebenenfalls mit schwer verständlichen Abkürzungen unterhalb der X-Achse zufrieden geben.

Mit *Punkten* können schließlich in einem zweidimensionalen Koordinatensystem die Werte zweier metrischer Merkmale gemeinsam dargestellt werden. Jeder Punkt symbolisiert eine Wertepaar (y_i, x_i) , wobei die $i = 1, \dots, n$ Wertepaare eine Punktwolke in einem sogenannten *Streudiagramm* bilden. Streudiagramme werden in dem Abschnitt 5.6.1 zur Darstellung des Zusammenhangs zweier metrischer Merkmale verwendet (Stichwort: bivariate Korrelation). Im Rahmen der Regressionsanalyse wird durch die Punktwolke eine Gerade gelegt, um die funktionale Abhängigkeit einer metrischen abhängigen Variablen Y von einer metrischen unabhängigen Variablen X zu quantifizieren (s. Abschnitt 5.6.2). Die Abbildung 3.8 illustriert ein fiktives Zahlenbeispiel für die gemeinsame Darstellung von 14 Wertepaaren $(y_i, x_i; i = 1, \dots, 14)$ der Merkmale Y und X. Graphikunterprogramme können die mathematisch zu bestimmende Gerade in eine empirische Punktwolke legen. Auf diese Form der graphischen Darstellung gehen wir ausführlich in den Abschnitten zur Korrelations- und Regressionsanalyse ein.

Die übliche graphische Darstellungsweise der Häufigkeiten von klassifizierten Daten ist das *Histogramm*. Hier ist es wichtig, daß die Klassen zunächst aneinanderstoßend gewählt werden. Bei einem Histogramm werden dann geschlossene Blöcke über die Klassen $x_j^U - x_j^O (j = 1, \dots, m)$ eingezeichnet. Sie repräsentieren die absoluten oder relativen Häufigkeiten der Klassen. Die Flächen dieser Blöcke beschreiben jeweils die Häufigkeit einer Klasse, d.h. nicht die Höhe des Blocks über einer Klasse ist proportional zur Klassenhäufigkeit, sondern der Flächeninhalt der Blöcke wird hier proportional zur Häufigkeit gewählt. Somit ist das Produkt von Blockhöhe und Klassenbreite einer Klasse proportional zur beobachteten Klassenhäufigkeit (Prinzip der Flächentreue). Auf der Ordinate (Höhe) wird somit $h_j / (x_j^O - x_j^U)$ über der j-ten Klasse abgetragen. Diese Blockhöhen werden auch als Häufigkeitsdichten bezeichnet oder kurz: Dichten. Histogramme werden ebenso wie Säulendiagramme erstellt, um die Häufigkeitsstruktur der Daten herauszustellen. Bei der graphischen Darstellung der Häufigkeitsverteilung können, wenn es sich um Klassen mit konstanter Breite handelt, in dem Histogramm die Klassenhäufigkeiten direkt als Blockhöhen verwendet werden. Die Abbildung 3.9 zeigt ein fiktives Zahlenbeispiel für Klassenhäufigkeiten, wobei von konstanten Klassenbreiten $x_j^U - x_j^O (j = 1, \dots, 5)$ ausgegangen wird.

In der Abbildung 3.9 ist neben dem Histogramm auch ein sogenanntes *Häufigkeitspolygon* eingetragen, ein Liniendiagramm, das alternativ zum Histogramm die Häufigkeitsverteilung darstellt. Für die Konstruktion des aus dem Histogramm hervorgehenden Polygons benötigt man zunächst die Klassenmitten M_j , wobei generell $M_j = 1/2(x_j^O - x_j^U)$ ist. Der Polygonzug ist dann die Verbindungsstrecke der jeweils zu den Klassenmitten gehörenden Punkten auf den Oberkanten der Säulen (in der Abbildung die fett ausgezogene Linie). Bei

i	x_i	y_i
1	3	11
2	4	14
3	5	13
4	5	16
5	6	20
6	8	21
7	10	22
8	11	26
9	12	24
10	12	29
11	13	28
12	16	31
13	17	34
14	18	33

Wertepaare $y_i, x_i, i = 1, \dots, 14$ in einem Streudiagramm:

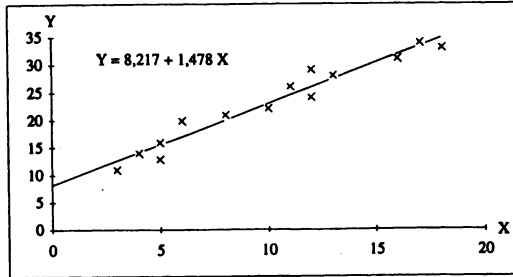


Abb. 3.8: Fiktive Wertetabelle zweier metrischer Merkmale Y und X, dargestellt als zweidimensionale Punktwolke in einem Streudiagramm mit einer Geradenfunktion

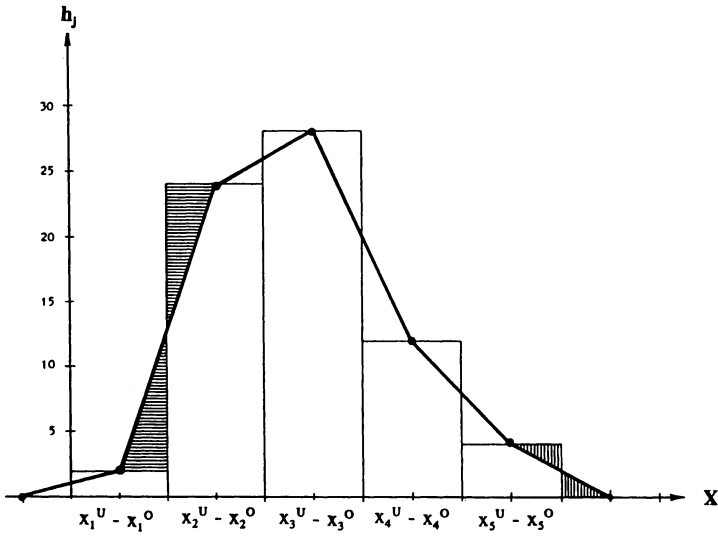


Abb. 3.9: Histogramm mit Häufigkeitspolygon

Quelle: Bohley, P., 1992: Statistik. Einführendes Lehrbuch für Wirtschafts- und Sozialwissenschaftler, 5. überarb. u. erw. A., München, S. 93.

dieser Form der Darstellung wird in der Regel das Histogramm vernachlässigt. Das Häufigkeitspolygon -auch als *Häufigkeitsdichtekurve* bezeichnet – vermittelt ebenso wie ein Histogramm einen optischen Eindruck der Häufigkeitsverteilung. Es läßt sich zeigen, daß die Flächen unter dem Histogramm und unter dem Polygon gleich groß sind, wenn man an beiden Enden der Verteilung jeweils eine halbe (fiktive) Klassenbreite berücksichtigt. Die jeweils am rechten und linken Ende abgeschnittenen Dreiecke der Rechtecksflächen entsprechen den fiktiven halben Endklassen. Der Nachteil ist allerdings, daß diese fiktiven Klassen kaum zu interpretieren sind und die gesamte Verteilung künstlich gestreckt wird, wodurch der graphische Gesamteindruck leicht verzerrt wird.

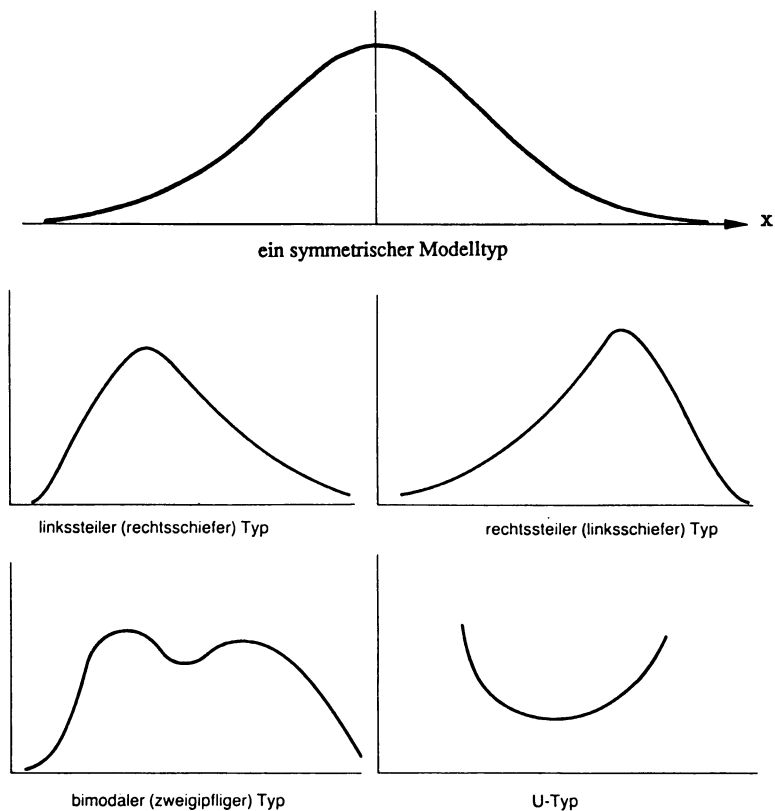


Abb. 3.10: Exemplarische Modelltypen für Häufigkeitsverteilungen

Quelle: Bohley, P., 1992: Statistik. Einführendes Lehrbuch für Wirtschafts- und Sozialwissenschaftler, 5. überarb. u. erw. A., München, S. 94f.

Für den statistischen Laien ist die Darstellung des Polygons weniger verständlich, da sie wegen des Bezuges zu den künstlich geschaffenen Klassenmitten

nur mit zusätzlichen Annahmen interpretierbar ist. Allerdings kann das Häufigkeitspolygon bei einer großen Anzahl von Klassen die Bestimmung typischer Verteilungsformen erleichtern. Der Polygonzug geht mit zunehmender Klassenzahl und abnehmender Klassenbreite mehr und mehr in einen gerundeten Kurvenzug über. Es gibt verschiedene Typen solcher gerundeter Kurvenzüge, die in der Abbildung 3.10 exemplarisch als fünf Modelltypen repräsentiert sind.

Man kann für eine graphische Darstellung neben den absoluten oder relativen Häufigkeiten auch kumulierte (absolute oder relative) Häufigkeiten verwenden, um die (absoluten oder relativen) Summenhäufigkeiten graphisch zu veranschaulichen. Eine solche Darstellung heißt *Summenhäufigkeitsfunktion* für kumulierte absolute Häufigkeiten bzw. Prozentanteile oder *empirische Verteilungsfunktion* $F(x)$ mit einem Wertebereich zwischen 0 und 1, wenn die aufwärtskumulierten relativen Häufigkeiten zugrundegelegt werden.

Die Summenfunktion ist eine Alternative zum Häufigkeitspolygon und gibt die graphische Antwort auf die Fragen: Wieviel Prozent der Datenwerte sind *kleiner oder gleich* $X = x_j$ oder wieviel Prozent sind *gleich oder größer* $X = x_j$. Für die graphische Darstellung wird daher zunächst eine tabellarische Aufwärtssummierung bzw. Abwärtssummierung der ursprünglichen Häufigkeiten der Merkmalsausprägungen $x_j^0 - x_j^o$ ($j = 0, 1, \dots, m$) vorgenommen.

a) Die absolute Summenhäufigkeit $H_j(x_j)$ der j -ten Merkmalsausprägung x_j^o ist:

$$H_j(x_j^o) = 0 + h_1 + \dots + h_j.$$

b) Die relative Summenhäufigkeit $P_j(x_j^o)$ der j -ten Merkmalsausprägung x_j^o ist:

$$F_j(x_j^o) = 0 + h_1 / n + \dots + h_j / n.$$

Die Konstruktion eines Polygonzuges mit kumulierten Häufigkeiten geht von der Annahme aus, daß die einzelnen Untersuchungseinheiten innerhalb der Klassen gleichverteilt sind. Folglich steigt der Graph der Verteilungsfunktion von der Klassenuntergrenze bis zur Klassenobergrenze linear an. Erst bei Erreichen der Obergrenze der jeweiligen Klasse wird der Wert der kumulierten Häufigkeit erreicht. Bei der Konstruktion der Summenkurven dürfen daher auch nicht die Klassenmitten verbunden werden! Vielmehr werden jetzt die Klassenobergrenzen miteinander verbunden (Ausnahme: Beginn der Kurve bei einem fiktiven Startwert von Null). Eine Summenhäufigkeitsfunktion geht ebenfalls bei abnehmender Klassenbreite und entsprechender Zunahme der Zahl der Klassen in einen gerundeten Kurvenzug über.

Die Tabelle 3.5a zeigt am Beispiel einer Aufwärtskumulierung die allgemeine Berechnung von absoluten und prozentualen relativen Summenhäufigkeiten, die Tabelle 3.5b illustriert die Berechnung anhand des oben verwendeten Zahlenbeispiels. Die graphischen Darstellungen zu Tabelle 3.5b zeigt die Abbildung 3.11. Zur Veranschaulichung der Konstruktion sind die Häufigkeiten je Klasse als gestrichelte Rechtecke eingezeichnet. Ebenso wie bei stati-

Tab. 3.5a: Allgemeine Tabelle für eine Aufwärtskumulierung von absoluten und relativen Häufigkeiten

Größenklasse K_j von dem Merkmal X_k	absolute Häufigkeit h_j	absolute Summen- häufigkeiten	relative Summen- häufigkeiten
kleiner als x_1^U	0	0	0.0
$x_1^U - x_1^O$	h_1	h_1	h_1/n
$x_2^U - x_2^O$	h_2	$h_1 + h_2$	$h_1/n + h_2/n$
$x_3^U - x_3^O$	h_3	$h_1 + h_2 + h_3$	$h_1/n + h_2/n + h_3/n$
.	.	.	.
.	.	.	.
.	.	.	.
$x_j^U - x_j^O$	h_j	$h_1 + h_2 + \dots + h_j$	$h_1/n + \dots + h_j/n$
.	.	.	.
.	.	.	.
.	.	.	.
$x_m^U - x_m^O$	h_m	$h_1 + \dots + h_j + \dots + h_m$	$h_1/n + \dots + h_j/n + \dots + h_m/n$
Summenzeile:	n	n	1.0

Tab. 3.5b: Zahlenbeispiel für aufwärts- und abwärtskumulierte Prozentanteile

Klassen K_j von X_k	absolute Häufigkeiten h_j	Prozentanteile $(h_j/n)*100$	cum. h_j in Prozent	decum. h_j in Prozent
unter x_1^U	0	0 %	0,00 %	100,00 %
$x_1^U - x_1^O$	2	3 %	2,86 %	97,14 %
$x_2^U - x_2^O$	24	34 %	37,14 %	62,86 %
$x_3^U - x_3^O$	28	40 %	77,14 %	22,86
$x_4^U - x_4^O$	12	17 %	94,29 %	5,71
$x_5^U - x_5^O$	4	6 %	100,00 %	0.00 %
Summe:	70	100 %	-	-

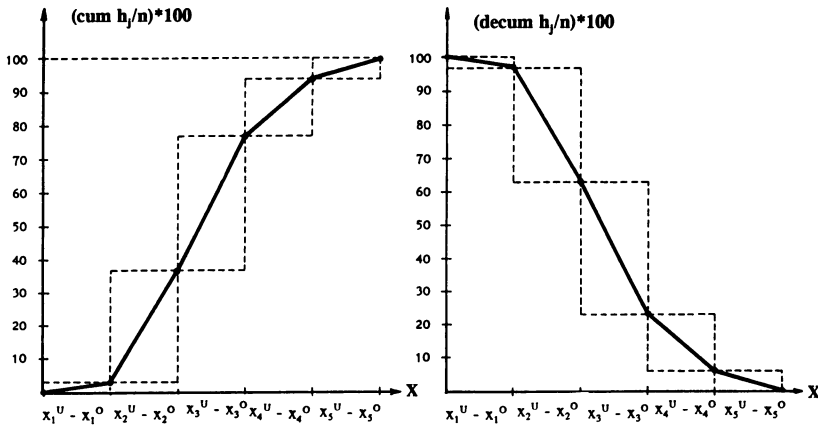


Abb. 3.11: Aufwärts- und abwärtskumulierte Prozentanteile

Quelle: Bohley, P., 1992: Statistik. Einführendes Lehrbuch für Wirtschafts- und Sozialwissenschaftler, 5. überarb. u. erw. A., München, S. 96.

stischen Tabellen ist es erforderlich, daß der Betrachter einer Graphik deren Inhalt und Aussage schnell erfassen kann. Folgende Elemente gehören daher zu einer vollständigen Graphik:

- Titel: Überschrift für die Graphik.
- Untertitel: Eine zusammenfassende Aussage oder Kurzbeschreibung der Daten nach Raum und Zeit.
- Fußnote: Ergänzende Angaben zum Inhalt oder zu Details der Graphik.
- Quellenangaben: Herkunft der Daten.
- Achsenbeschriftung: X-Achse (Beschreibung der horizontalen Achse), Y-Achse (Beschriftung der senkrechten Achse).
- Skala mit Unterteilung: Kurzer Teilstrich, der einen Wert auf der Skala der Y- und X-Achse markiert und Bezeichnung der Maßeinheit.
- Legende: Erläuterung der unterschiedlichen Schraffuren (bzw. Graustufen) einzelner Flächen oder einzelner Symbole o.ä..

Neben den Inhalten einer Graphik gibt es noch die äußeren Gestaltungsmerkmale (Layout). Diese können einen erheblichen Einfluß auf die Lesbarkeit der Informationen ausüben, die mit Hilfe der Graphik vermittelt werden. Ohne hierauf näher einzugehen, sind folgende Aspekte zu nennen: Die Verwendung von Farben oder Graustufen in der Graphik, die Verwendung von Schraffuren oder Raster, die Wahl eines geeigneten Formats, die Wahl der Schriftart und Schriftgröße, die Verwendung von Bildsymbolen, Hintergrundzeichnungen bei der graphischen Ausgestaltung von statistischen Diagrammen.

Abschließend stellen wir die wichtigsten Vorteile graphischer Darstellungen zusammenfassend heraus (ausführlich s. Geßler, J.R., 1993: Statistische Graphik, Basel u.a., S. 17ff):

- Eine Graphik hat auf ihren Betrachter eine stimulierende Wirkung. Sie dient innerhalb von Texten als Blickfang für den Leser. Durch das Betrachten einer Graphik hat der Leser die Möglichkeit, die wesentlichsten Aussagen schnell zu erfassen. Eine Tabelle ist zwar ebenfalls ein Blickfang, hat aber den Nachteil, die Informationen nicht in so globaler Weise vermitteln zu können. Das gilt insbesondere dann, wenn es sich um relativ große und unübersichtliche Tabellen handelt.
- Die Informationsaufnahme erfolgt durch Graphiken wesentlich schneller als durch eine tabellarische Darstellung. Der Leser benötigt weniger Augenbewegungen. Eine Graphik wird schließlich ganzheitlich wahrgenommen, bei einer Tabelle registriert das Auge Details, z.B. einzelne Werte. Auf graphischen Wege kann wesentlich mehr Information aufgenommen werden. Die graphische Information wird als gesamte Informationseinheit wahrgenommen und im Gedächtnis gespeichert. Diese Informationseinheiten können mit weniger gedanklichem Aufwand verarbeitet werden.
- Eine graphische Darstellung ist für Menschen im Gedächtnis länger verfügbar als Zahlendarstellungen. Ferner ist das bildliche Gedächtnis des Menschen um ein vielfaches leistungsfähiger, was die Verarbeitung und Speicherung von Informationen betrifft.
- Eine graphische Darstellung liefert eine einfache Beschreibung der tabellierten statistischen Ergebnisse. Sie vermittelt selbst den statistisch nicht vorgebildeten Betrachter einen Gesamteindruck von den Zahlenwerten. Der Betrachter kann sich leicht ein anschauliches Bild von den statistischen Ergebnissen machen und diese im Gedächtnis speichern. In tabellarischen Darstellungen tritt die besondere Gestalt eines Ergebnisausschnitts im allgemeinen nicht, oder zumindest nicht deutlich genug, hervor.
- Eine Graphik ist nicht nur eine beschreibende Darstellung, sondern sie hat gegenüber einer Tabelle häufig auch eine den Betrachter aktivierende Wirkung. Die graphische Darstellung macht deutlicher auf die Struktur der Ergebnisse aufmerksam. Damit wird die zentrale Aussage der Ausgangstabelle in den Vordergrund gerückt.
- Graphiken sind anschaulicher als numerische Informationen. Durch diese Eigenschaft können Sachverhalte lebendiger dargestellt werden.

4. Zur Anwendung statistischer Modelle in der empirischen Sozialforschung

4.1 Einleitung: Zum Stellenwert statistischer Modelle

Dem Begriff *Statistik* wird gewöhnlich eine zweifache Bedeutung unterlegt: Zum einen wird Statistik als *Methodenlehre* für den Umgang mit quantitativen Daten aufgefaßt, zum anderen versteht man darunter auch die Ergebnisse der Anwendung dieser Methodenlehre, z.B. die zusammenfassende Darstellung des Datenmaterials in Form von Tabellen, Graphiken oder statistischen Kennzahlen. Als Wissenschaft (*theoretische Statistik*) ist Statistik ein Zweig der Mathematik und primär nicht durch ihr Anwendungsgebiet, sondern durch die statistischen Modelle und Verfahrensweisen (Methoden) gekennzeichnet. Die Modelle, die die statistische Methodenlehre zur Verfügung stellt, können zur Beantwortung solcher Fragestellungen eingesetzt werden, deren Gegenstandsbereich durch quantifizierbare Merkmale beschrieben werden kann. Nicht jede Analyse von Massenerscheinungen kann unmittelbar als Statistik bezeichnet werden. Erst durch die Anwendung formaler (d.h. mathematischer) Modellvorstellungen wird die Analyse von Massenerscheinungen zur Statistik.

Der Wert der Statistik für die historische Sozialforschung besteht darin, daß sie formale Methoden bereitstellt, mit deren Hilfe aus der unüberschaubaren Fülle empirischer Tatsachen die im Zusammenhang mit einer bestimmten Fragestellung relevanten Informationen herausgefiltert werden können. »Die von der Statistik als Wissenschaft zur Verfügung gestellten Modelle und Methoden sind für den empirisch arbeitenden Sozialwissenschaftler ein wichtiges Hilfsmittel, um die in den Daten enthaltenen Informationen durch geeignete Transformationen des Zahlenmaterials herauszuarbeiten und zu verdichten, so daß diese Informationen – die empirischen Ergebnisse – für die Ziele der Untersuchung verwertet werden können (*angewandte Statistik*). Die Statistik kann dabei *nicht mehr* sein als ein Hilfsmittel; sie kann keine Informationen *produzieren*, die nicht schon – wenn auch in weniger deutlicher Form – in den Daten vorhanden sind« (Kromrey, H., 1991: Empirische Sozialforschung, Opladen, S. 312). Die Abbildung 4.1 verdeutlicht die Stellung der Statistik im Forschungsprozeß.

Immer wenn Modelle – seien sie, mathematisch betrachtet, sehr elementar oder sehr komplex – mit realen Beobachtungsdaten konfrontiert werden, kommt die Statistik zum Einsatz. In der Statistik gibt es verschiedene Modelltypen. Die sog. *deskriptive Statistik* beschreibt konkret vorliegende Daten, ohne daß Verallgemeinerungen über den Stichprobendatensatz hinaus vorgenommen werden. Die Modelle der deskriptiven Statistik helfen, numerische Daten zu ordnen (z.B. in Form von einfachen Häufigkeitstabellen) und grafisch zu veranschaulichen (z.B. in Form von Linien-, Balken- oder Kreisdiagrammen). Hinzu kommt die Verwendung von statistischen Kennwerten zur (komprimierten) Kennzeichnung der drei wichtigsten Aspekte von Merkmalen: die zentrale

Inhaltliche Ebene

Statistisch-methodische Ebene

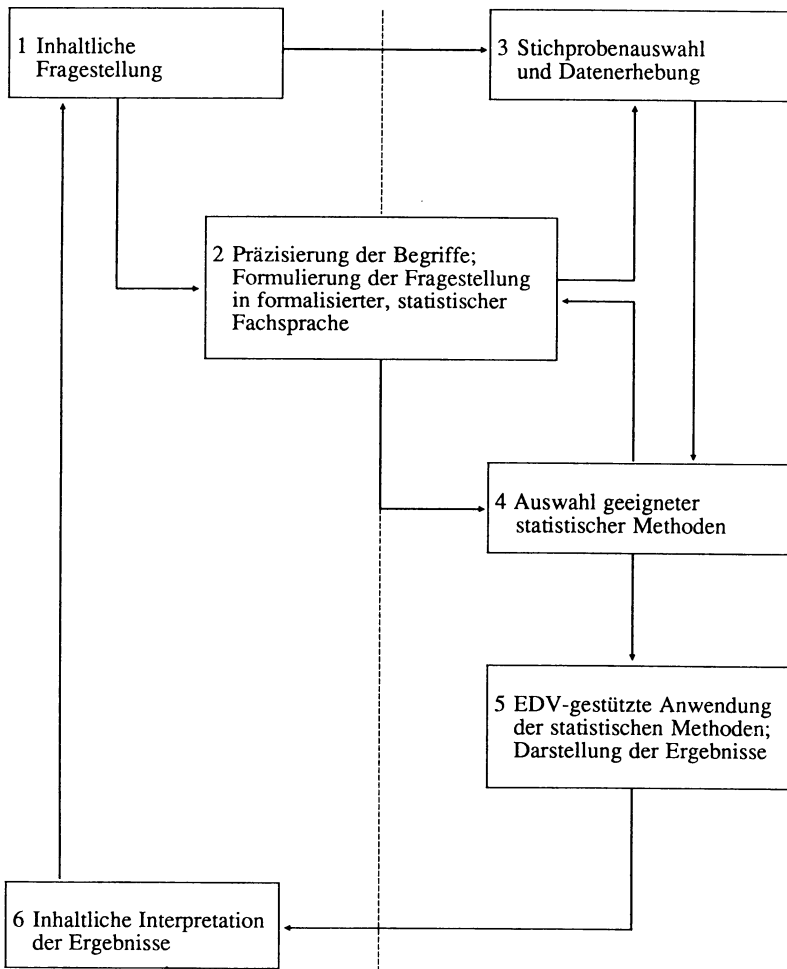


Abb. 4.1: Die Stellung der Statistik in der historischen Sozialforschung

Tendenz, die Streuung univariater Verteilungen und der Zusammenhang zwischen zwei simultan betrachteten Merkmalsverteilungen. Behandelt werden hier vor allem methodisch einfachere Probleme wie die Berechnung von Mittelwerten und Streuungsmaßen sowie die Berechnung von Maßzahlen zur Kennzeichnung der Zusammenhangsstärke zweier Merkmale.

Ferner sind Modelle entwickelt worden, die Verallgemeinerungen von Stichprobenergebnissen auf Grundgesamtheiten ermöglichen. Damit ist das Gebiet der wahrscheinlichkeitstheoretisch fundierten sog. schließenden oder *induktiven Statistik* (Inferenzstatistik) angesprochen. Sehr vereinfacht kann sie als Brücke zwischen der beschreibenden Statistik und der Wahrscheinlichkeitstheorie aufgefaßt werden. Sie umfaßt Problemkreise wie statistisches Schätzen und statistische Tests.

Darstellungen der beschreibenden Statistik in statistischen Lehrbüchern sind oft insofern irreführend, als daß der Eindruck erweckt wird, die beschreibende Statistik hätte nur die Verwendung von »einfachen« statistischen Modellen zum Gegenstand. Die Modelle sind nur insofern »einfach«, als daß in ihnen die Wahrscheinlichkeitsrechnung im inferenzstatistischen Sinn keine Rolle spielt. Der entscheidende Unterschied zur schließenden Statistik liegt nicht in der vermeintlich geringeren Vielfalt der Verfahren, sondern in der Perspektive (Schluß von einer Stichprobe auf die Grundgesamtheit).

Werden simultan mehrere Merkmale in die statistische Analyse einbezogen, gewinnen Modelle an Bedeutung, die Beziehungen zwischen mehreren Merkmalen repräsentieren. Diese Modelle fallen in das weite Gebiet der sogenannten *multivariaten Statistik*. Im Kern lassen die (komplexeren) Verfahren der multivariaten Statistik ebenfalls die beiden Perspektiven »Deskription« und »statistische Inferenz« zu.

Entscheidend für den Einsatz von Statistik im Forschungsprozeß ist die angemessene Wahl eines Modells bei gegebener Fragestellung und Datenkonstellation (Meßniveaus der erhobenen Merkmale). Nicht gerade selten ist eine fehlerhafte oder falsche Anwendung statistischer Modelle bei Fragestellungen zu beobachten, für die diese Modelle nicht entwickelt worden sind. Unangemessene Modelle liefern keine gültigen Ergebnisse.

Mit der Herausstellung des Begriffes *Modell* im Rahmen der Statistik verfolgen wir ein bestimmtes Ziel. Wenn z.B. in den Sozialwissenschaften von Modellen die Rede ist, dann werden sehr unterschiedliche Typen und Anwendungen angesprochen. Zahlreiche Buchtitel verweisen direkt auf den Aspekt der Modellbildung, wie ein Blick in die Bibliographie zeigt:

- »Causal models in the social sciences« (Blalock, H.M., Chicago 1971),
- »Theorie und Modell« (Ziegler, R., München 1972),
- »Modelling binary data« (Collett, D., London 1991),
- »Social dynamics: Models and methods« (Tuma, N.B./Hannan, M.T., New York 1984).

- »Data Modelling – Modelling History« (Titel der internationalen Konferenz der »Association for History & Computing«, August 1996).

Diese Liste ließe sich beliebig fortführen. Modellhaftes Denken oder die Anwendung von Modellen impliziert zunächst eine »natürliche« Tendenz in Richtung Formalisierung. Zum anderen verhilft modellhaftes Denken zu der wichtigen Unterscheidung zwischen dem empirischen Datenmaterial und den Modellen, mit denen dieses (Roh-) Material konfrontiert wird. Ein Modell ist eine stark verkürzte Abbildung möglicher (sozialer) Strukturen und Prozesse. Die Verkürzung ist zweckgebunden, sie entsteht aus dem Erkenntnisinteresse der wissenschaftlichen Disziplin oder des einzelnen Forschers. Ein Modell ist somit ausschnitthaft und mit dem Untersuchungsgegenstand nicht identisch; Ziel ist Homomorphie (Strukturähnlichkeit) zwischen Modell und Realität, im Idealfall Isomorphie (Strukturgleichheit). In beiden Fällen geht es aber nur um Struktur, nicht um vollständige Identität.

In einem ganz konkreten Zusammenhang spricht Jürgen Kriz einen Aspekt dieser Unterscheidung an, wenn er dafür plädiert, nicht von statistischen Methoden, sondern von statistischen Modellen zu sprechen (vgl. Kriz, J.: 1973: Statistik in den Sozialwissenschaften, Reinbek bei Hamburg, S. 23f). Eine Methode oder ein Verfahren verwendet man, um von einem genau definierten Anfangszustand zu einem definierten Endzustand zu gelangen³. Entscheidend ist, daß unterschiedliche Methoden, auf identische Ausgangsdaten angewandt, zu genau demselben Ziel führen: Der Endzustand ist methodenunabhängig. Bei einem Modell dagegen handelt es sich um ein Abbild einer definierten Ausgangsstruktur unter bestimmten, angebbaren Gesichtswinkeln⁴. Verschiedene Gesichtswinkel sind denkbar, je nach Erkenntnisinteressen des Betrachters, aber jede Betrachtungsweise stellt nur einen bestimmten Gesichtswinkel dar.

Die Statistik bietet exklusiv die Möglichkeit an, wissenschaftliche Aussagen mit erhobenem Datenmaterial in Beziehung zu bringen und diese Relation zu bewerten. Erst sie macht aus Wissenschaften auf elaboriertem Niveau Erfahrungswissenschaften. Es ist aber von zentraler Bedeutung, daß eine rationale Wahl bezüglich des Elaboriertheitsgrades anzuwendender Modelle getroffen werden kann. Das heißt u.a.: das elaborierteste Modell braucht nicht das adäquateste Modell zu sein. Kriterien dafür, ob ein statistisches Modell der zu untersuchenden Fragestellung adäquat ist oder nicht, lassen sich nicht aus der

³ So gibt es verschiedene Methoden, um aus 12 Zahlen etwa das arithmetische Mittel zu berechnen: z.B. alle 12 Zahlen zu addieren, dann durch 12 dividieren; oder jede Zahl zunächst durch 12 dividieren, dann die Ergebnisse addieren

⁴ Als Beispiel kann der Stadtplan als Modell einer Stadt gelten. Man kann ein solches Modell unter verschiedenen Gesichtspunkten erstellen: als maßstabgetreue Straßenkarte oder als Touristenkarte mit Hervorhebungen der Sehenswürdigkeiten. Hier wird nicht die physische Totalität abgebildet, sondern nur die Lage der Straßenzüge, die es ermöglichen, einen bestimmten Ort zu finden. Man kann auch ein dreidimensionales, maßstabgetreues Modell der Stadt erstellen. Jedes dieser Modelle ist für andere Zwecke brauchbar.

Statistik allein beziehen, sondern müssen vor dem Hintergrund der Hypothesenbildung und Operationalisierung entwickelt werden. »Die Brauchbarkeit der einzelnen Modelle hängt ausschließlich davon ab, wieweit sie der (sozial-) wissenschaftlichen Fragestellung angemessen sind – und nicht von ihrem Stellenwert innerhalb der (mathematischen) Statistik-Theorie« (Kriz, a.a.O., S. 14). Der Modellcharakter der Statistik betont, daß eine Begründung geleistet werden muß, warum überhaupt innerhalb der Sozialwissenschaften bestimmte statistische Modelle angewandt werden und welche Funktionen die einzelnen Modelle in diesem Zusammenhang haben.

Die Entscheidung für ein bestimmtes statistisches Modell setzt auch voraus, daß die Modellbedingungen die Bedingungen des Realitätsausschnitts vollständig reproduzieren oder – wenn die Bedingungen nur unvollständig repräsentiert sind – ob beobachtbare Abweichungen der Modellannahme von dem Realitätsausschnitt noch tolerierbar sind, ohne dadurch die Gültigkeit der Anwendung zu gefährden. Im Sinne modellhaften Denkens können die Relationen zwischen Theorie, Hypothesen und Daten einerseits und induktivem versus deduktivem Vorgehen andererseits betrachtet werden (s. Abbildung 4.2).

Auf der *deduktiven Seite* wird davon ausgegangen, daß man aus Theorien Hypothesen ableiten kann. Um sie einer empirischen Überprüfung zugänglich zu machen, sind sie in ein Modell zu übersetzen, das die Hypothesenstruktur adäquat repräsentiert. Die Statistik konfrontiert dann das Modell mit den Beobachtungsdaten. Falls die Hypothesen nicht haltbar sind, bedeutet dies, daß entweder die Theorie nicht zutrifft, daß die Theorie für den konkreten Anwendungsbereich unzulässig ist oder daß die von der Theorie postulierten Hypothesen nicht adäquat in ein Modell umgesetzt worden sind.

Auf der *induktiven Seite* verfügt der Forscher über Daten, die etwas mit seinem spezifischen Gegenstandsbereich zu tun haben. Die Statistik unterstützt in diesem Fall datengenerierte Theoriebildungsversuche. Zu den modellhaften Resultaten werden dann Erklärungen gesucht, die zu theoretischen Sätzen umgeformt werden. Diese Forschungsperspektive ist allerdings lediglich von beschreibender Natur, die per se keine Verallgemeinerungen zulassen. Für beide Perspektiven gilt: Die Modellarbeit interveniert in das Daten-Theorie-Verhältnis (s. Abbildung 4.3).

Wir haben es in der auf statistische Daten gegründeten historischen Sozialforschung stets mit drei begrifflichen Ebenen zu tun, einer theoretischen, einer statistischen und einer den Beobachtungsbefund beschreibenden Ebene. Zwischen ihnen bestehen enge Beziehungen, aber auch grundsätzliche Unterschiede. Die Theorie-Sprache beherrscht das sozialwissenschaftliche Denken. Sie bildet (allgemein) die Referenz für eine konkrete empirische Untersuchung. Die Beobachtungsbegriffe schaffen der Theorie den Zugang zur Realität. Für sie allein erhalten wir quantitative Daten. Statistische Modelle sind jeweils durch ihre spezifische Modellcharakteristik (Modelltheorie) geprägt, die einen spezifischen Zugang zu dem betrachteten Realitätsausschnitt schafft, der in den

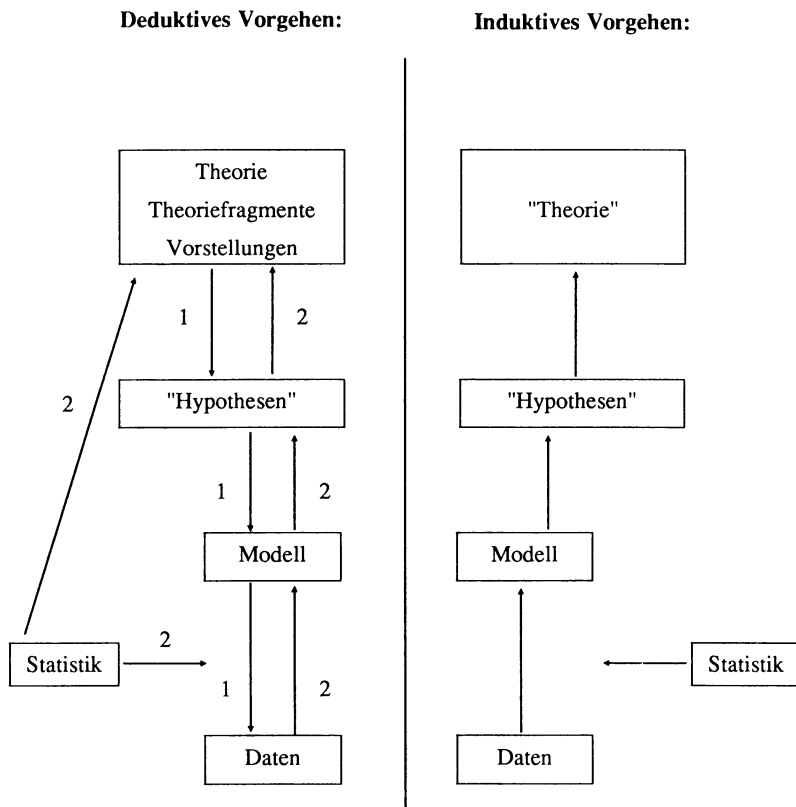


Abb. 4.2: Deduktives versus induktives Vorgehen

beobachteten Daten repräsentiert ist. Die Modelltheorie »stellt sich dem Problem, wie man die einander fremden Welten in Korrespondenz zueinander bringen kann« (vgl. Harder, T., 1975: Daten und Theorie, München, S. 56f). »Der agonistische Stil des Daten-Theorie-Verhältnisses beruht darauf, daß beide von Haus aus einander fremd sind und nur durch einen partiellen Souveränitätsverzicht die eine mit der anderen Seite in Berührung kommen kann. Ein solcher Verzicht wird durch die Einsicht nahegelegt, daß eine Konfrontation zwischen theorieempfindlichen Daten und daten-kommensurabler Theorie nur auf dem beiden verhaßten, aber unumgänglichen neutralen Boden eines Modells stattfinden kann« (Harder, a.a.O., S. 141).

Die hier diskutierte Unterscheidung ist auf die Betrachtung von Theorie (oder Hypothesen), Untersuchungsdesign, Merkmale, Daten und dem Modell der statistischen Konfrontation von Theorie und Daten im sozialwissenschaft-

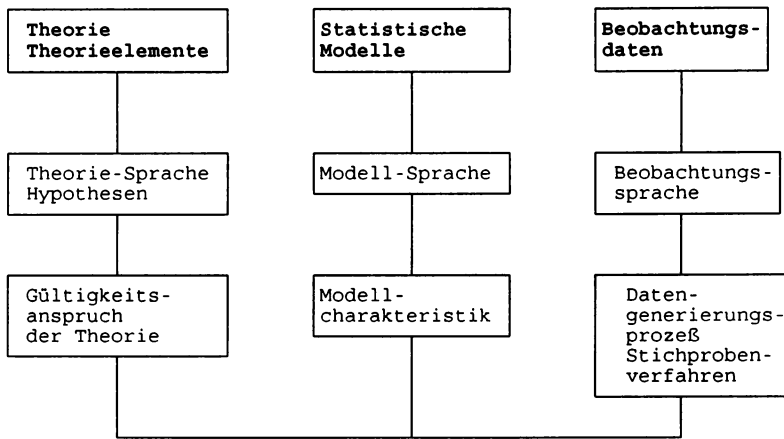


Abb. 4.3: Verhältnis von Daten-Theorie-Modell

lichen Forschungsprozeß auszudehnen (s. Abbildung 4.4). Aus dem zentralen Stellenwert statistischer Modellarbeit lassen sich fünf grundlegende Anforderungen an statistische Modelle ableiten:

1. Ein statistisches Modell sollte dem Datentyp gerecht werden; das Meßniveau der Merkmale wird nicht künstlich verändert, da sonst ein Informationsverlust vorliegt.
2. Die Theorie sollte erkennbar sein; Merkmale sind theoretisch begründet zueinander in Beziehung zu setzen (in der bivariaten bzw. multivariaten Datenanalyse).
3. Das statistische Modell sollte für einen multivariaten Ansatz offen sein, in dem simultan mehr als zwei Merkmale in die Datenanalyse einbezogen werden.
4. Das statistische Modell sollte dem Untersuchungs-/Erhebungsdesign gerecht werden (z.B. Querschnitts- versus Längsschnittdaten, Zeitreihen usw.).
5. Aus dem statistischen Modell sollten informationsverdichtende Maßzahlen ableitbar sein.

Die in der theoretischen Statistik entwickelten Modelle beruhen auf idealisierten Annahmen über Eigenschaften von Daten, insbesondere über Meßniveaus und Verteilungsformen. Eine entscheidende Aufgabe ist deshalb vor der Anwendung der Statistik immer die Prüfung, ob die statistischen Modellannahmen bei gegebener Datenbasis erfüllt sind oder ob sie zumindest hinreichend genau erfüllt sind. Bei der Auswahl des statistischen Modells ist daher zweierlei zu

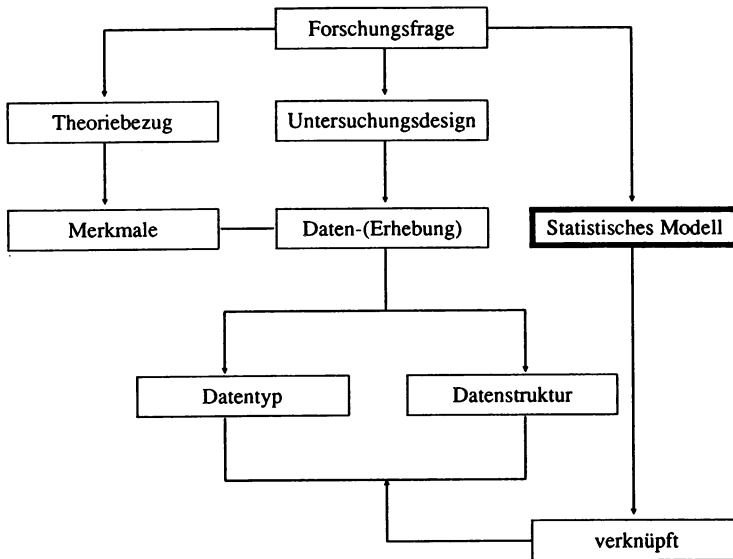


Abb. 4.4: Stellenwert statistischer Modelle

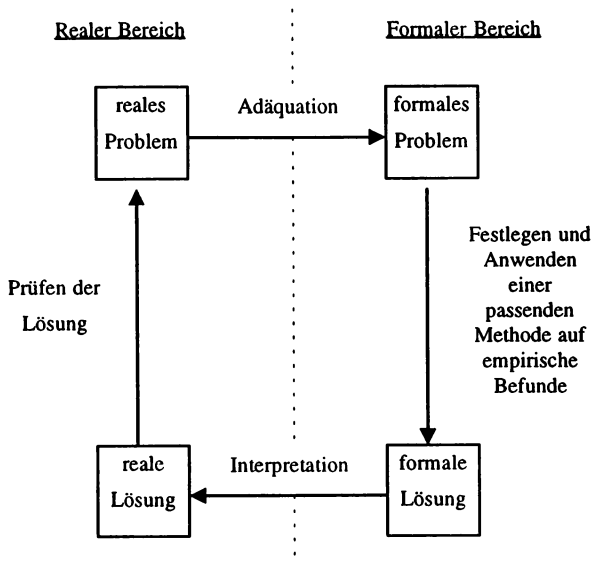


Abb. 4.5: Prozeß der empirischen Erkenntnisbildung

beachten: (1) Die Statistik stellt für unterschiedliche Zwecke verschiedene mathematische Modelle der Informationsverarbeitung zur Verfügung, mit unterschiedlichen Stärken, Schwächen und Reichweiten bezüglich der ableitbaren substantiellen Aussagen. Unter diesem Modellspektrum sind (bei gegebener Datenstruktur und gegebenen Meßniveaus der Merkmale) die für die eigenen Untersuchungsziele sachgerechtesten statistischen Modelle auszuwählen. Für ein und denselben Zweck stellt aber die Statistik oft mehrere dem Datenniveau angepaßte Modelle zur Verfügung. (2) Es ist daher unter den verschiedenen, für ein und denselben Zweck angelegten Modellen eines auszuwählen, das dem Datenniveau angeglichen ist.

In der statistischen Wissenschaft werden fortgesetzt Datenanalysemodelle weiter- oder neuentwickelt. Sie entfaltet dabei eine Tendenz zur Eigendynamik, bei der oft die Strenge der mathematischen Argumentationen mehr gilt als deren praktische Verwertbarkeit. Die sozialwissenschaftliche Forschungspraxis neigt dazu, diese Problematik zu überschätzen und ist deswegen oftmals skeptisch gegenüber neueren statistischen Modellentwicklungen, insbesondere dann, wenn diese in einer mathematisch anspruchsvollen Form präsentiert werden. Die Statistik neigt im Gegensatz eher dazu, die Anwendungsproblematik zu unterschätzen, weil sie oftmals ein allzu idealisiertes Bild von der Realität hat. Zur Veranschaulichung dieses Problemkreises dient die Abbildung 4.5. »Dieses Kreisdiagramm verdeutlicht den – leider häufig vernachlässigten – Unterschied zwischen der Realität (z.B. der ökonomischen oder der sozialen Sphäre) und der formalen Untersuchung der Realität mit Hilfe von statistischen Methoden. Das Schema weist ferner darauf hin, daß die empirische Erkenntnis zwei Transponierungsprozeduren erfordert: Die Adäquation, d.h. die Anpassung des formalen Problems an das reale Problem, und die Interpretation der formalen Lösung in der realen Sphäre. Ohne Einsicht in diese Zusammenhänge kann die Statistik nicht ihre eigentliche Aufgabe erfüllen, realwissenschaftliche Probleme sachgerecht, wirkungsvoll und in nachprüfbarer Weise zu durchleuchten. Eine Beschränkung des Gesichtsfeldes auf die formalen Aspekte führt fast zwangsläufig dazu, die Statistik entweder nur wie einen Zweig der abstrakten Mathematik zu behandeln oder aber als Methode zum Erzeugen von Zahlenfriedhöfen anzusehen« (Schäffer, K.-A., 1980: Zur Entwicklung der statistischen Methodik und ihrer Anwendungen, in: Allgemeines Statistisches Archiv, 64, S. 1f). In Anlehnung an *Heinrich Hartwig*⁵ bezeichnet Schäffer den Transformationsprozeß zwischen einem realen und formalen Problem als *Adäquation*. Adäquation bedeutet hier u.a. auch die Modellierung der Abhängigkeit z.B. zwischen zwei Merkmalen als Voraussetzung für die Auswahl und Anwendung eines statistischen Modells. Die mit Hilfe eines Modells erreichbare Lösung ist dann jedoch zunächst eine formale; sie besteht aus Tabellen, Grafiken, statistischen Kennzahlen oder in Form einer Regressionsfunktion etc.

⁵ Siehe Hartwig, H., 1956: Naturwissenschaftliche und sozialwissenschaftliche Statistik, in: Zeitschrift für die gesamte Staatswissenschaft, 112, S. 262.

Die formale Lösung verlangt daher am Ende noch eine Rücktransformation in den realen Bereich. Dies geschieht durch Interpretation als einer zusätzlichen Leistung, die zur realen Lösung führt. Der formale Bereich bildet das zentrale Betätigungsfeld der Statistik.

Die Auswahl eines dem realen Problem adäquaten Modells aus der Menge der inzwischen verfügbaren Datenanalysemodelle (s. Teil 3: Auswahlbibliographie) erfordert zwar einige Mühe und viel Fingerspitzengefühl, ist aber unabdingbare Voraussetzung für eine sachgerechte Analyse des Ausgangsproblems. Der wesentliche Fortschritt in der Entwicklung statistischer Modelle ist darin zu sehen, daß inzwischen für unterschiedliche Untersuchungsdesigns und Merkmale mit beliebigen Skalenniveaus sehr leistungsfähige Datenanalysemodelle zur Verfügung stehen. Der verhältnismäßig große Rechenaufwand, den viele der neu entwickelten Modelle erfordern, stellt kein Hindernis für ihre Anwendung dar, da der Einsatz von Statistik-Programmpaketen die Anwendung eines breiten Modellspektrums ermöglicht.

4.2 Grundlegende Konzepte für das Verständnis statistischer Modelle auf deskriptiver Ebene

Die in dem vorangegangenen Abschnitt dargestellten tabellarischen und graphischen Häufigkeitsverteilungen stellen bereits eine erste Zusammenfassung oder Verdichtung der ursprünglich beobachteten Einzeldaten (d.h. des Urmaterials) dar. Die Auswertung des statistischen Datenmaterials ist selten auf diese Verdichtungsformen beschränkt. Oft ist man auch noch daran interessiert, Häufigkeitsverteilungen in knapper Form zu charakterisieren. Eine wichtige Modellklasse der univariaten Datenanalyse bildet die Charakterisierung von Häufigkeitsverteilungen durch einige wenige statistische Maßzahlen oder Stichproben-Kenngrößen (bzw. Parameter im Fall der Grundgesamtheit), die in komprimierter Form die wesentlichen Eigenarten univariater Verteilungen widerspiegeln. Diese Maßzahlen sollen z.B. die *Lage* der Verteilung (ihre *zentrale Tendenz*), ihre Form (*Schiefe* und *Wölbung*) oder die *Variabilität der Einzelwerte* um das Zentrum beschreiben (*Streuung*).

4.2.1 Lagemaße

Lagemaße geben Auskunft über das Zentrum (die Lage) einer Häufigkeitsverteilung. Diese Kennzahlen werden auch als *Mittelwerte* bezeichnet. Die wichtigsten *lagetypischen Mittelwerte* sind der Modus und der Median.

Modus: Die geringsten Anforderungen an das Skalenniveau stellt der Modus oder Modalwert (M). Der Anwendungsbereich des Modus liegt dort, wo nach dem typischen Wert einer Häufigkeitsverteilung gefragt wird. Unter dem Modus versteht man den in einer Verteilung am häufigsten vorkommenden Merkmalswert (auch: dichtester Wert). Je nachdem ob ein diskretes oder ein in

Klassen gruppiertes (ursprünglich metrisch gemessenes) Merkmal vorliegt, ermittelt man den Modus aus der Häufigkeitstabelle als die Merkmalsausprägung x_j mit der größten absoluten Häufigkeit h_j bzw. im Falle klassifizierter Merkmale die Klassenmitte jener Klasse K_j mit der größten absoluten Häufigkeit (im Falle gleichbreiter Klassen). Der Modus ist auch aus einem Säulendiagramm oder Histogramm an dem höchsten Abstand zur x-Achse erkennbar.

Ermittlung des Modus aus der Häufigkeitstabelle:

$$M = x_j^* \quad (\text{diskretes Merkmal})$$

x_j^* ist die Merkmalsausprägung mit der maximalen Häufigkeit.

$$M = \frac{1}{2} \cdot (x_j^{\circ*} + x_j^{u*}) \quad (\text{stetiges Merkmal})$$

$x_j^{\circ*}$ und x_j^{u*} sind die Klassengrenzen der Klasse K_j^* mit der maximalen Häufigkeit.

Er kann somit nur dann ermittelt werden, wenn eine Häufigkeitsverteilung vorliegt, bei der eine Merkmalsausprägung eine größere Häufigkeit aufweist als alle anderen Ausprägungen. Der Modus kann bei Daten jeden Meßniveaus bestimmt werden und ist sehr einfach zu interpretieren. In der Tabelle 3.3b (Gemeindegrößenklassen in der Weimarer Republik) ist der Modus die Mitte der zweiten Klasse (3.000 Einwohner). Dieses Lagemaß vermittelt allerdings nur einen sehr geringen Informationsgehalt. Hat die Häufigkeitsverteilung sehr viele Ausprägungen und sind die Häufigkeiten je Ausprägung entsprechend klein, dann ist die Verwendung dieser Maßzahl nicht mehr sinnvoll.

Der Modus enthält relativ wenig Informationen. Es besagt nur: der Modalwert kommt häufiger vor als jeder andere Wert der Verteilung. Wenn die übrigen Werte aber fast ebenso häufig vorkommen wie der Modus, kann ein Wert mehr oder weniger aus Zufall zum Modus werden. Nur wenn ein Wert eindeutig dominiert, sollte man auf den Modus zurückgreifen. Besitzt eine Häufigkeitsverteilung eines ordinalen oder metrischen Merkmals einen unteren oder oberen Extremwert, dann ist der Modus kein befriedigendes Maß für die zentrale Tendenz.

Median: Als Lagemaßzahl wird der Median auch als *Zentralwert (Z)* oder 50-Prozentpunkt bezeichnet. Voraussetzung für die Ermittlung des Medians ist eine Ordnung der empirisch ermittelten Merkmalswerte nach ihrer Größe (Bildung einer Rangwertreihe). Als Lagemaßzahl ist der Median daher erst ab ordinalem Skalenniveau eines Merkmals sinnvoll. Der Zentralwert bezeichnet dann denjenigen Merkmalswert, der in der Mitte der Meßwertreihe liegt. Er teilt somit die Beobachtungseinheiten in zwei gleich große Gruppen. Links

und rechts des Medians liegen genau je 50% der nach Größe geordneten Meßwertreihe. Für die Bestimmung des Medians ist zu unterscheiden, ob die Anzahl der vorliegenden Datenwerte n gerade oder ungerade ist. Die Definition führt zu folgenden Berechnungsformeln für den Median aus der Rangwertreihe:

Ermittlung des Medians aus der Rangwertreihe:

$$Z = \begin{cases} \frac{1}{2} (x_{n/2} + x_{(n/2)+1}) & \text{für } n \text{ gerade} \\ x_{(n+1)/2} & \text{für } n \text{ ungerade} \end{cases}$$

x_i sind die neu indizierten Merkmalswerte aus der Rangwertreihe.

Bei einem ungeraden n existiert ein mittleres Element, so daß der Zentralwert auf einfache Weise abzulesen ist. Für die geordnete Folge von X-Werten

1 ; 1 ; 4 ; 5 ; 7

gilt somit $Z = 4$. Bei geradem n existiert kein mittleres Element. Der Median ergibt sich in diesem Fall dadurch, daß der Mittelwert der beiden Werte gebildet wird, die innerhalb der Rangreihe an den Positionen » $n/2$ « und » $n/2+1$ « platziert sind. Für die Folge von geordneten X-Werten

1 ; 1 ; 4 ; 5 ; 7 ; 9

gilt somit $Z = (4 + 5)/2 = 4,5$.

Bei größeren Fallzahlen ist eine größenmäßige Auflistung sehr umständlich. Hier empfiehlt es sich zunächst eine kumulierte (Summen-) Häufigkeitsverteilung zu erstellen, aus der das Ergebnis dann leicht abgelesen werden kann. Der Zentralwert entspricht genau der Ausprägung x_j , bei der die kumulierte relative Häufigkeit den Wert 0.5 erstmals überschreitet.

Schwieriger ist die Berechnung des Medians bei klassifizierten Beobachtungswerten. In diesem Fall kann man den Median nicht mehr durch Abzählen oder durch Rückgriff auf die geordnete Häufigkeitsverteilung ermitteln. In diesem Fall ist eine Schätzung notwendig. Zunächst ist die Klasse zu bestimmen, in der die Summenhäufigkeitsfunktion den Wert 0.5 erreicht (Medianklasse); anschließend wird durch Interpolation der Zentralwert näherungsweise bestimmt:

$$Z = \text{Untergrenze der Medianklasse} + \left[\frac{n/2 - \left(\text{kumulierte Häufigkeit unterhalb der Medianklasse} \right)}{\left(\text{Häufigkeit innerhalb der Medianklasse} \right)} \right] \cdot \text{Breite des Intervalls}$$

Der Median hat die Eigenschaft, unempfindlich gegenüber einzelnen extremen Merkmalswerten zu sein (robustes Lagemaß). Dies ist von Bedeutung, da bei empirischen Daten *Ausreißer* recht häufig vorkommen. In diesem Fall möchte man verhindern, daß dadurch die Beschreibung der zentralen Lage des Merkmals beeinflußt wird.

Die *rechnerischen Mittelwerte* – *arithmetisches* und *geometrisches Mittel* – erhält man durch algebraische Berechnung. Wir beschränken die Darstellung im folgenden auf das arithmetische Mittel.

Arithmetisches Mittel: Das wohl am häufigsten verwendete Lagemaß ist das arithmetische Mittel (Mittelwert, Durchschnitt), das nur für metrisch skalierte Merkmale verwendet werden darf. Im Gegensatz zu den Kennzahlen Modus und Zentralwert ist das arithmetische Mittel eine Maßzahl, bei dessen Berechnung jeder einzelne Merkmalswert berücksichtigt wird. Liegen Einzelmeßwerte vor, dann ist das arithmetische Mittel definiert als die durch den Umfang der statistischen Erhebung n dividierte Summe aller Merkmalswerte x_i ($i=1, \dots, n$).

Ermittlung des arithmetischen Mittels aus der Urliste:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Das arithmetische Mittel ist auch derjenige Wert einer nach ihrer Größe geordneten Reihe von X -Werten, von dem aus die Summe der Abweichungen nach oben und nach unten gleich groß ist, d.h. die Abweichungen vom arithmetischen Mittel (unter Berücksichtigung der Vorzeichen) addieren sich zu Null auf:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Diese Eigenschaft stellt seine primäre Legitimation dar. Man sagt auch: Das arithmetische Mittel ist der *Schwerpunkt* einer Verteilung (analog zum physikalisch verstandenen *Massenmittelpunkt*). Wählt man als Abstandsmaß die mittleren quadrierten Abweichungen, so erfüllt das arithmetische Mittel sogar die sogenannte Kleinstquadrateneigenschaft (s.a. Abschnitt 5.6.2 zum Schätzprinzip der linearen Einfachregression), d.h.: die Summe der quadrierten Abweichungen der Beobachtungswerte zum Mittelwert ergibt ein Minimum (Minimum-Eigenschaft). Die Minimum-Eigenschaft des Mittelwerts läßt sich wie

folgt interpretieren: Soll unter Unkenntnis des jeweils beobachteten Wertes x_i ($i=1, \dots, n$) für jede Untersuchungseinheit die Merkmalsausprägung durch einen einheitlich gewählten Prognosewert vorhergesagt werden, so daß die Summe der quadratischen Abweichungen von den beobachteten Werten insgesamt am kleinsten ist, so stellt der Mittelwert von X den besten Prognosewert dar.

$$\sum_{i=1}^n (x_i - \bar{x})^2 \Rightarrow \min$$

Das arithmetische Mittel wertet auch die Abstände zwischen den Beobachtungswerten aus. Für den Median dagegen ist es unerheblich, wie sich die Beobachtungswerte links und rechts von ihm verteilen; es entscheidet allein die Tatsache, daß links und rechts von ihm 50% der Beobachtungswerte liegen. Bei dem arithmetischen Mittel dagegen bekommen – durch die Einbeziehung aller Einzelwerte – die extremen Werte einer Verteilung ein hohes Gewicht. Das arithmetische Mittel ist somit ausreißeranfällig (d.h. diese Maßzahl ist nicht robust).

Bei der Bestimmung des arithmetischen Mittels aus bereits tabellierten Daten ergeben sich leichte Änderungen gegenüber dem Vorgehen, das durch die Definition oben nahegelegt wird. Liegt eine Häufigkeitstabelle ohne Klassifizierung vor, so können die Daten in eine geordnete Form gebracht werden. Sind x_1, \dots, x_m die empirischen Ausprägungen des Merkmals X und h_1, \dots, h_m die entsprechenden absoluten Häufigkeiten, dann folgt die Berechnung des Mittelwertes aus folgender Gleichung:

Ermittlung des arithmetischen Mittels aus der Häufigkeitstabelle:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^m h_j x_j$$

Da die einzelnen Merkmalsausprägungen in der Häufigkeitsverteilung mit den absoluten Häufigkeiten gewichtet oder *gewogen* werden, bezeichnet man den Durchschnitt in diesem Fall auch als *gewogenes Mittel*.

Im Fall klassifizierter Daten kann das arithmetische Mittel nur näherungsweise bestimmt werden. Die genaue Lage der Werte in den jeweiligen Intervallen ist ja nicht bekannt. Man unterstellt daher, daß die Einzelwerte sich jeweils gleichmäßig innerhalb einer Klasse verteilen. Bei der Berechnung geht man in diesem Fall von den Klassenmitten $(x_j^o - x_j^u)/2$ als Mittelpunkt des Wertintervalls aus und berechnet ein gewogenes Mittel. Die Formel für das arithmetische Mittel aus der Häufigkeitstabelle kann somit bei klassifizierten Daten ebenfalls angewandt werden, wenn für x_j ($j = 1, \dots, m$) jeweils die Klassenmitte eingesetzt wird.

Tab. 4.1: Median und Mittelwert zu ausgewählten Strukturmerkmalen auf Wahlkreisebene

Strukturmerkmal	Median	arithm. Mittel
Prozentsatz der römisch-katholischen Bevölkerung	16	30,5
Prozentsatz der evangelischen Bevölkerung	71	64,4
Prozentsatz der Bevölkerung in Gemeinden mit bis zu 5.000 Einwohnern	52	44,2
Prozentsatz der Arbeiter und und erwerbslosen Arbeiter	39	39,3
Prozentsatz der Arbeitslosen an allen Erwerbspersonen	18	18,5

Lesart: Der erste Medianwert »16%« teilt die 35 Wahlkreise hinsichtlich des Merkmals »Prozentsatz der römisch-katholischen Bevölkerung« in zwei gleich große Hälften auf, d.h. 50% der Wahlkreise weisen Werte kleiner/gleich 16% auf und die restlichen 50% der Wahlkreise weisen Werte größer 16% auf (der $(35 + 1)/2 = 18$. Wert ist gleich 16). Der durchschnittliche Prozentanteil der römisch-katholischen Bevölkerung beträgt (über 35 Wahlkreise gemittelt) 30,5%; die durchschnittliche prozentuale Arbeitslosenquote 18,5%.

Als empirisches Beispiel legen wir im folgenden die Datenmatrix aus der historischen Wahlforschung mit ausgewählten Sozialstrukturdaten auf der Untersuchungsebene der Wahlkreise in der Weimarer Republik zugrunde (s. Tab. 3.1b im letzten Abschnitt). Die Tabelle 4.1 enthält die Mediane und Mittelwerte zu ausgewählten Sozialstrukturmerkmalen. Die behandelten drei Werte zielen darauf ab, die gesamte Information innerhalb einer Häufigkeitsverteilung auf einen einzigen *typischen* Wert zu reduzieren. Für das statistische Modell *Modus* heißt »typisch«: »am häufigsten vorkommend«, für das Modell *Median*: »in der Mitte einer geordneten Reihe liegend«, für das Modell *arithmetisches Mittel*: »den Schwerpunkt der gesamten Beobachtungswerte darstellend«. Die Tabelle 4.2 und die Abbildung 4.6 stellen die wichtigsten Lagemaße noch einmal zusammenfassend gegenüber. Die drei diskutierten Maße der zentralen Tendenz einer univariaten Verteilung lassen sich in ihrem Verhältnis zueinander wie

Tab. 4.2: Mögliche Lagemaße und ihr Anwendungsspektrum

Skalenniveau des Merkmals X	Mittelwert	Zusätzlicher Informations- gehalt
nominal- skaliert	Modus	Unterscheidung
ordinal- skaliert	Modus Median	Unterscheidung Rangfolge
intervall- skaliert	Modus Median arithm. Mittel	Unterscheidung Rangfolge Abstände

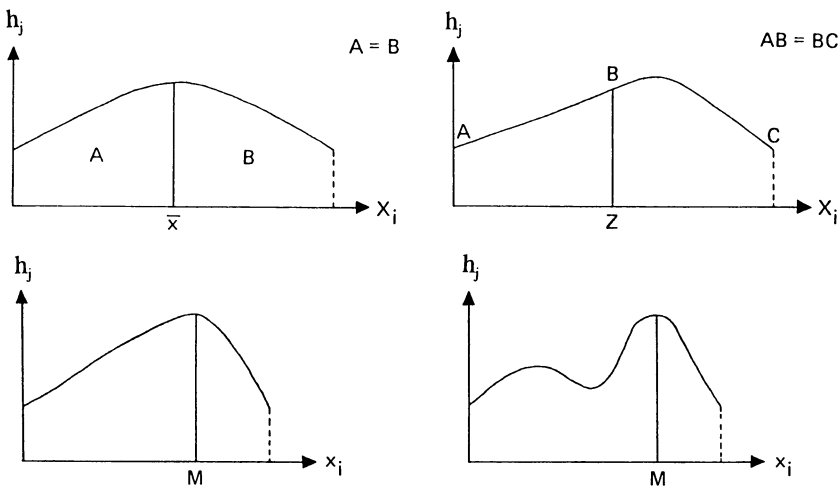


Abb. 4.6: Grafische Veranschaulichung der Lagemaße

folgt charakterisieren (vgl. Kromrey, H., 1991: Empirische Sozialforschung, 5., überarb. u. erw. A., Opladen, S. 338):

- »Bei Daten, aus denen ein arithmetisches Mittel berechnet werden darf (die also mindestens intervallskaliert sind), können auch der Median (Zentralwert) und der Modus bestimmt werden. Die drei Maße stellen jedoch jeweils eine andere Information in den Vordergrund und weisen daher normalerweise unterschiedliche Zahlenwerte auf. Die Entscheidung für eines dieser Maße muß also von der Fragestellung der Auswertung her begründet werden.

- Lediglich im Sonderfall einer eingipfligen und vollkommen symmetrischen Verteilung fallen das arithmetische Mittel, der Zentralwert und der Modus zusammen.
- Bei asymmetrischen eingipfligen Verteilungen liegt der Modus immer genau unter dem Gipfelpunkt der Kurve; Median und arithmetisches Mittel liegen auf der flacheren Seite, wobei das arithmetische Mittel (da es von den extremen Werten der Verteilung stark beeinflusst wird) den größeren Abstand vom Modus aufweist.«

Statistisch interessant wird demnach die Beziehung zwischen den Maßzahlen der zentralen Tendenz dann, wenn man die empirisch ermittelten Werte für das arithmetische Mittel, den Median und den Modus vergleicht. Das Verhältnis dieser drei Mittelwerte zueinander erlaubt Rückschlüsse auf die Häufigkeitsverteilung, aus der sie stammen. Die Abbildung 4.7 stellt eine graphische Lagebestimmung der drei Mittelwerte für unterschiedliche Verteilungen dar.

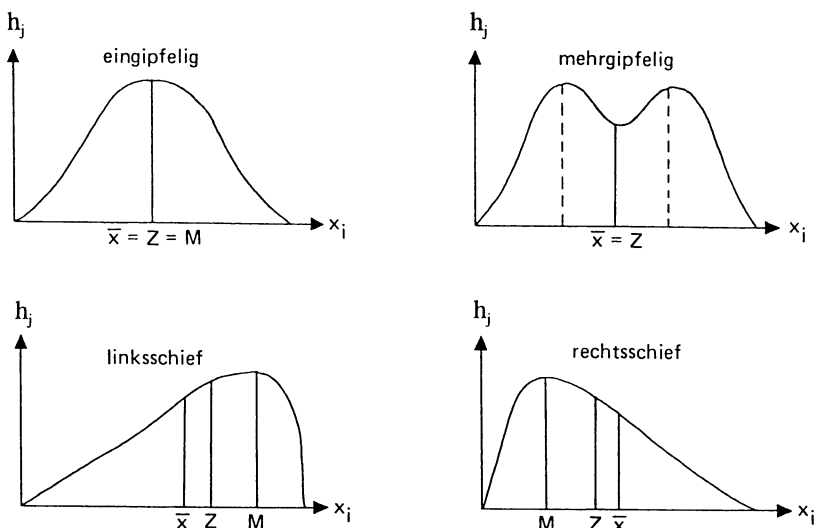


Abb. 4.7: Das Verhältnis der Lagemaße zueinander für verschiedene Häufigkeitsverteilungen

Die *Schiefe* ist neben der zentralen Tendenz ein weiteres wichtiges Einzelcharakteristikum einer Häufigkeitsverteilung. Eine Möglichkeit zur Beurteilung der Schiefe einer Verteilung bietet der Vergleich von arithmetischem Mittel, Modus und Median (*Fechnersche Lageregel*) als eine grobe Basis zur Einstufung von Verteilungen.

Bei einer eingipfligen, symmetrischen Verteilung fallen das arithmetische Mittel, der Zentralwert und der häufigste Wert zusammen (Mittelwert = Median = Modus). Stimmen diese Werte nicht überein, so ist dies ein Anzeichen

dafür, daß die beobachteten Daten eine schiefe, asymmetrische Verteilung besitzen. *Rechtsschiefe* Verteilungen zeichnen sich dadurch aus, daß die meisten Merkmalsausprägungen kleine und mittlere Werte aufweisen, während große Werte relativ selten auftreten. Eine rechtsschiefe Verteilung hat ihr arithmetisches Mittel oberhalb des Zentralwertes und des Modus (arithmetisches Mittel $>$ Median $>$ Modalwert). Bei *linksschiefen* Verteilungen liegen diese Lagemaßzahlen gerade in umgekehrter Größenordnung vor: Eine linksschiefe Verteilung hat ihren Schwerpunkt (arithmetisches Mittel) unterhalb des Zentralwertes und des Modus (arithmetisches Mittel $<$ Median $<$ Modalwert).

Betrachten wir noch einmal zwei Ergebnisse in Tabelle 4.1. Das Merkmal »Prozentanteil der Arbeitslosen« ist symmetrisch verteilt. Eine deutlich rechtsschiefe Verteilung zeigt sich dagegen bei dem Merkmal »Prozentsatz der römisch-katholischen Bevölkerung«.

Mittelwerte informieren zwar über die zentrale Tendenz einer Verteilung, sie geben allerdings keinen Aufschluß über die Homogenität bzw. Heterogenität einer Häufigkeitsverteilung. Unter letzterem Gesichtspunkt sind im folgenden Abschnitt noch die Modelle zur Quantifizierung der *Streuung* zu besprechen.

4.2.2 Streuungsmaße

Die bloße Angabe einer Lagemaßzahl beschreibt eine Häufigkeitsverteilung nur unzureichend. Sie liefert zwar die zentrale Tendenz einer univariaten Verteilung, gibt aber keine Auskunft darüber, wie weit die Beobachtungswerte innerhalb einer Stichprobe (oder einer Grundgesamtheit) streuen. Neben der Lage ist die *Streuung* (Variabilität, Variation, engl.: dispersion) das wichtigste Charakteristikum einer Verteilung. Sollen Aussagen über das Vorliegen einer *Homogenität* (Gleichartigkeit) der Untersuchungseinheiten bezüglich eines Merkmals X getroffen werden bzw. die Art der Heterogenität (Verschiedenartigkeit) beschrieben werden, so ist die Variabilität um das Zentrum ein geeignetes Maß. Für eine statistische Maßzahl, die eine Aussage über den Grad der jeweiligen Variabilität macht, sollte daher folgendes Prinzip zutreffen: je kleiner der Wert dieser Maßzahl ist, desto häufiger treten Werte auf, die relativ nahe am Zentrum der Verteilung liegen, d.h. desto mehr konzentrieren sich diese Werte um das Zentrum. Die Erfassung der Streuung ist an das metrische Skalenniveau der Merkmale gebunden, denn nur für diesen Skalentyp ist ein Abstand zwischen Merkmalswerten definiert. Die Streuungsmaße (Dispersionskoeffizienten) dienen zur Charakterisierung der Abstände zwischen den Beobachtungswerten eines Merkmals. Sie sind eine Ergänzung zur Mittelwertkennzahl und informieren auch darüber, welche Qualität der Mittelwert als Kennzahl der Verteilung hat. Streuungsmaße können auch als *Gütemaße* für Mittelwerte verstanden werden. Die Abbildung 4.8 veranschaulicht zwei Häufigkeitsverteilungen mit gleicher zentraler Tendenz, aber unterschiedlicher Streuung.

Die einfachste Kennzeichnung der Streuung eines metrischen Merkmals erfolgt durch die Angabe der *Spannweite* (Streuungsintervall, Variationsbreite).

Die Spannweite ist definiert als Differenz zwischen dem größten und kleinsten Beobachtungswert einer Verteilung.

Ermittlung der Spannweite aus der Urliste:

$$R(\text{ange}) = \max_i (x_i) - \min_i (x_i)$$

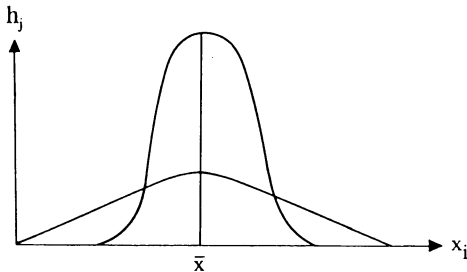


Abb. 4.8: Gleicher Mittelwert zweier Verteilungen bei unterschiedlicher Streuung

Aus einer Häufigkeitsverteilung erhält man die Spannweite bei m Merkmalsausprägungen als Differenz zwischen größter (x_m) und kleinster (x_1) Ausprägung. Als Beispiel betrachten wir das erste Merkmal aus der Datenmatrix der Sozialstrukturmerkmale auf Wahlkreisebene (s. Tabelle 3.1b): Der kleinste römisch-katholische Prozentanteil an der Bevölkerung beträgt 3% (Leipzig und Chemnitz-Zwickau), der größte Prozentanteil liegt bei 96% (Niederbayern). Die Spannweite beträgt somit 93 (Wertebereich der ersten Merkmals).

Liegt eine Häufigkeitsverteilung für ein klassifiziertes Merkmal vor, ist die Spannweite die Differenz zwischen der oberen Klassengrenze (x_m^U) und der unteren Klassengrenze (x_1^U).

Da die Spannweite lediglich von den Extremwerten bestimmt ist, kann sie nur als Angabe des Wertebereichs interpretiert werden. Insbesondere beim Vorliegen von extremen Werten (Ausreißern) ist dieses Streuungsmaß problematisch. Da es lediglich auf den beiden Extremwerten der Verteilung beruht, sagt es nichts über die Streuung der übrigen Werte aus, die zwischen den Extremwerten liegen.

Eine Alternative zur Spannweite zur Charakterisierung der Streuung einer Verteilung besteht in der Berechnung von *Quantilen*. Man kann eine Verteilung in beliebig viele gleichgroße Teilgesamtheiten zerlegen (Perzentile, Quintile, Quartile, allgemein: Quantile). Am häufigsten ist eine Zerlegung in vier Teilgesamtheiten (Quartile). Ein Streuungsmaß, das nicht so abhängig von Extremwerten ist, ist der *Quartilsabstand*. Erforderlich ist zunächst – analog zur Bestimmung des Medians – die Ordnung der Merkmalswerte der Größe nach. Während der Median eine geordnete Merkmalswertereihe in zwei gleich große

Teilgesamtheiten zerlegt, teilen die drei Quartile Q_1 (unteres Quartil), Q_2 (mittleres Quartil) und Q_3 (oberes Quartil) die Gesamtheit der n Merkmalswerte in vier gleich große Teilgesamtheiten, wobei der X -Wert für das zweite Quartil gleich dem Median ist. Unterhalb des ersten Quartils liegen 25% der Merkmalswerte, unterhalb des dritten Quartils 75% der Merkmalswerte. Der Quartilsabstand sagt dann aus, wie groß das Werteintervall ist, das 50% aller Untersuchungseinheiten umfaßt, und zwar unter der Bedingung, daß die jeweils 25% niedrigsten und die 25% höchsten Merkmalswerte unberücksichtigt bleiben. In der Praxis wird oft auch der mittlere Quartilsabstand als einfacher Durchschnitt verwendet.

$QA = x(Q_3) - x(Q_1) \quad \text{bzw.} \quad x_{0.75} - x_{0.25}$ $MQA = \frac{x(Q_3) - x(Q_1)}{2} \quad \text{bzw.} \quad \frac{x_{0.75} - x_{0.25}}{2}$
--

Bei der Berechnung von Quartilen wird entsprechend vorgegangen wie bei der Berechnung des Medians. Die folgenden Rechenbeispiele mit geradem n illustrieren die Berechnung des Quartilsabstandes von geordneten Merkmalswerten aus symmetrischen Verteilungen:

- a) 2, 4, 6, 8, 10, 12, 14, 16; (Median: 9, Mittelwert: 9, Range: 14).
- b) 2, 5, 7, 8, 10, 11, 13, 16; (Median: 9, Mittelwert: 9, Range: 14).
- c) 2, 7, 8, 8,5, 9,5, 10, 11, 16; (Median: 9, Mittelwert: 9, Range: 14).

Die drei Verteilungen weisen bei gleichem Schwerpunkt und Range unterschiedliche Streuungen auf! Zunächst ermittelt man $1/4 \cdot n$ und $3/4 \cdot n$, d.h. die Anzahl der Fälle, die unterhalb von Q_1 und Q_3 liegen und bestimmt dann die Merkmalswerte $x_{25\%}$ und $x_{75\%}$:

- a) $x(Q_1) = 5 = (4+6) / 2$; $x(Q_3) = 13 = (12 + 14) / 2$.
- b) $x(Q_1) = 6 = (5 + 7) / 2$; $x(Q_3) = 12 = (11 + 13) / 2$.
- c) $x(Q_1) = 7,5 = (7 + 8) / 2$; $x(Q_3) = 10,5 = (10 + 11) / 2$.

Hat man das erste und dritte Quartil festgestellt, kann man den Quartilsabstand als Streuungsmaß bestimmen:

- a) $QA = 13 - 5 = 8,0$;
- b) $QA = 12 - 6 = 6,0$;
- c) $QA = 10,5 - 7,5 = 3,0$.

Die abnehmenden Quartilsabstände verdeutlichen die verschieden starken Streuungen in den drei Verteilungen. Um die Quartile einer Häufigkeitsverteilung zu bestimmen, kann man auch von der Summenfunktion ausgehen. Von den Ordinatenwerten 25% und 75% ausgehend, sucht man die zugehörigen

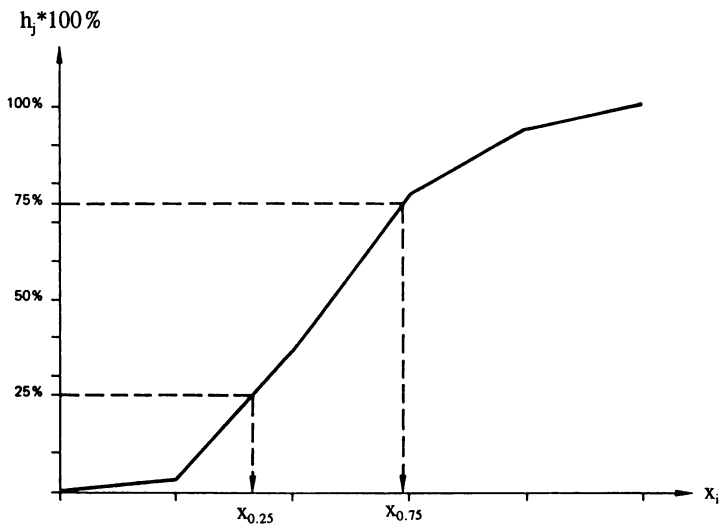


Abb. 4.9: Bestimmung des 1. und 3. Quartils aus einer Summenfunktion
 Quelle: Bohley, P., 1992: Statistik, München/Wien, S. 151.

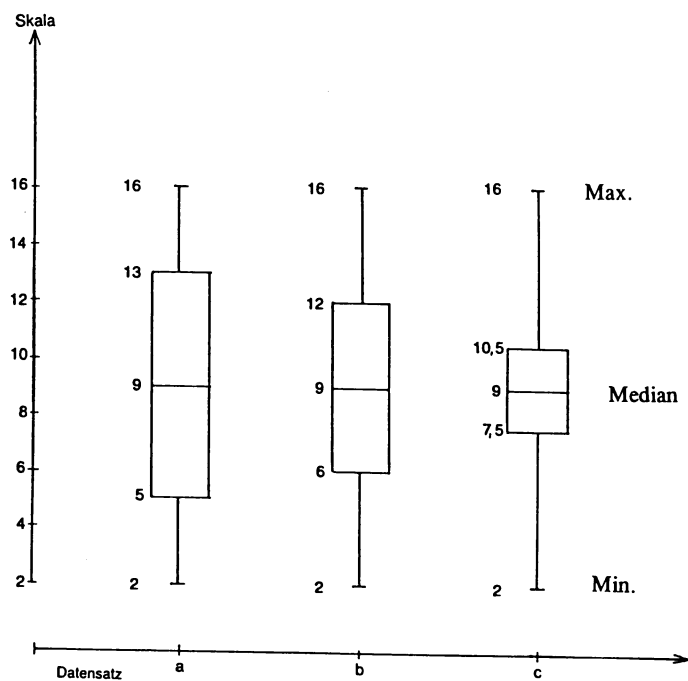


Abb. 4.10: Box-Diagramme im Vergleich
 Quelle: Bohley, P., 1992: Statistik, München/Wien, S. 153.

Werte auf der x-Achse auf. Die Abbildung 4.9 zeigt exemplarisch die Verwendung der Summenkurve zur Ermittlung des ersten und dritten Quartils für das Zahlenbeispiel in Tabelle 3.5b. Gegenüber der Spannweite hat der Quartilsabstand als Streuungsmaß den Vorteil, gegen Ausreißer unempfindlich zu sein. Allerdings beruht er analog zur Spannweite lediglich auf zwei Werten.

Zur graphischen Umsetzung der bisher dargestellten Streuungsmaße dienen *Schachtel-* oder *Box-Plots* (engl.: box-and-whisker plots). Die Streuung einer Verteilung wird damit einfach und anschaulich erfaßt. Bestandteile eines einfachen Box-Plots sind:

- Eine Skala parallel zur Hauptachse des Box-Plots.
- Ein Rechteck (Box), dessen Größe durch die Lage des ersten und dritten Quartils festgelegt wird. Die Breite der Box entspricht dem Quartilsabstand; sie ist somit der Bereich, in dem 50% der mittleren Merkmalswerte liegen.
- Der Median wird durch eine Linie innerhalb der Box angezeigt.
- Die Linien zwischen der Box und den Endpunkten werden *whisker* genannt. Die obere und untere Begrenzung der »whisker« – als Querstriche abgetragen – geben die Lage der Extremwerte einer Verteilung an.

Diese Darstellungsform ist insbesondere geeignet für den Vergleich verschiedener Datensätze (unterschiedliche Untersuchungen mit gleichem Merkmal X, Vergleiche im Zeitablauf oder Vergleich der Verteilung eines Merkmals X in verschiedenen Gruppen einer Stichprobe). Will man z.B. die Streuungen der Datenreihen (a), (b) und (c) graphisch veranschaulichen, ergeben sich die Box-Diagramme in der Abbildung 4.10. Mit Hilfe von Box-Plots können auch verschiedene Merkmalsverteilungen vergleichend dargestellt werden, sofern sie in den gleichen Dimensionen erhoben wurden. Als Beispiel betrachten wir drei Sozialstrukturmerkmale aus dem Wahlkreisdatensatz (s. Tab. 3.1b), die eine einheitliche Dimension aufweisen: Prozentanteil an der Bevölkerung des jeweiligen Wahlkreises. Die ausgewählten Merkmale weisen folgende Streuungskennwerte auf:

- Prozentanteile evangelische Bevölkerung: kleinster Wert = 4%; größter Wert = 95%; 1. Quartil = 49%; 3. Quartil = 86% (Median: 71%);
- Prozentanteile Bevölkerung in Gemeinden bis zu 5.000 Einwohnern: kleinster Wert = 0%; größter Wert = 80%; 1. Quartil = 32%; 3. Quartil = 58% (Median: 52%);
- Prozentanteile Arbeitslose an allen Erwerbspersonen: kleinster Wert = 9%; größter Wert = 35%; 1. Quartil = 13%; 3. Quartil = 23% (Median: 18%).

Diese Werte lassen sich anschaulich in nebeneinandergestellten Box-Plots graphisch veranschaulichen (s. Abbildung 4.11). Mittels dieser Box-Plots kann schnell ein Überblick über die Verschiedenartigkeit der Verteilungen hinsichtlich der Lage, der Streuung und der Schiefe, gewonnen werden. Ein Vergleich der Länge der »whisker« im Verhältnis zu der Länge des Rechtecks ist eine

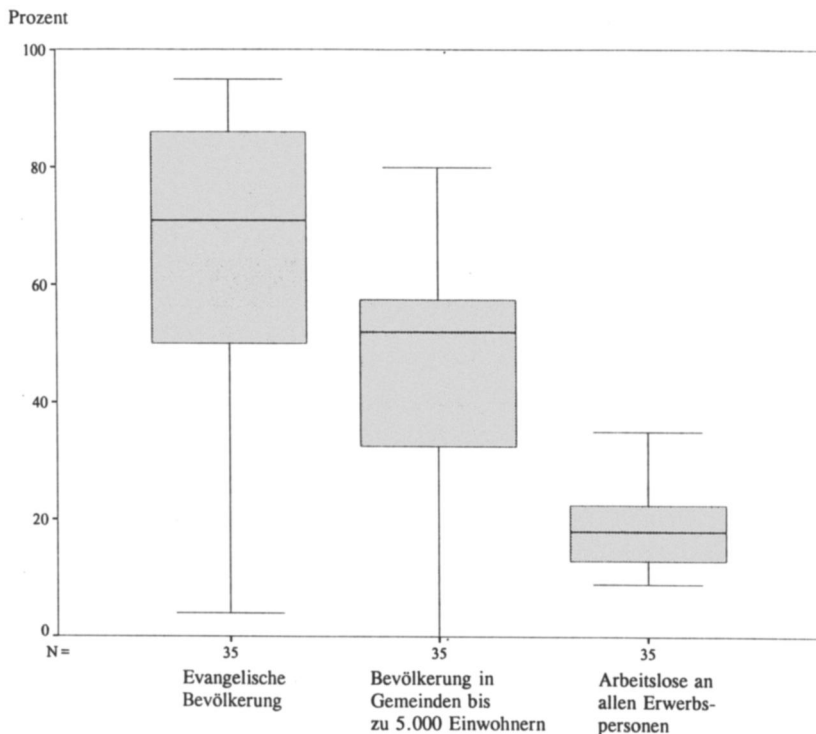


Abb. 4.11: Box-Plots zu ausgewählten Sozialstrukturmerkmalen auf Wahlkreisebene

Möglichkeit, die Streuung der Verteilungsenden zu beurteilen. Mit Hilfe der Box-Plots kann auch die Symmetrie der Datenverteilungen verglichen werden. Ist die Verteilung symmetrisch, so wird sich die gesamte Darstellung symmetrisch zum Median verhalten. Die Wahlkreise besitzen bezüglich der Arbeitslosenquote einen relativ niedrigen Schwerpunkt, eine deutlich niedrigere Streuung im Vergleich zu den übrigen Merkmalen des Diagramms und eine symmetrische Verteilung. Relativ symmetrisch verteilen sich auch die Prozentanteile der evangelischen Bevölkerung. Die Prozentanteile der Bevölkerung in Gemeinden bis zu 5.000 Einwohnern weisen eine stärker rechtsschiefe Verteilung auf. Sie haben eine deutlich geringere Streuung als die Prozentanteile der evangelischen Bevölkerung. Die unterschiedlichen Boxhöhen verdeutlichen sehr anschaulich die unterschiedlichen Lagen der drei Verteilungen.

Die Vorteile einer Box-Plot-Darstellung lassen sich in den folgenden Eigenschaften zusammenfassen.

- Die Mitte der Verteilung (genau 50%) wird betont.

- Extreme Beobachtungen können markiert und identifiziert werden.
- Der Quartilsabstand als einfaches Streuungsmaß kann leicht abgelesen werden.
- Box-Plots eignen sich gut für den Vergleich von Verteilungen.
- Die Schiefe der Verteilung ist optisch gut zu erkennen.

Die bisher behandelten Streuungskennzahlen berücksichtigen jeweils nur zwei Merkmalswerte in einer Beobachtungsreihe. Die im folgenden darzustellenden Streuungsmaße berücksichtigen sämtliche Beobachtungswerte einer Verteilung, indem sie eine durchschnittliche Abweichung der Merkmalswerte x_i ($i = 1, \dots, n$) von einem festen Wert messen. Da ein Streuungsmaß angeben soll, wie sehr die Beobachtungswerte vom Lagezentrum abweichen, ist es naheliegend, den mittleren Abstand der Daten vom Lagemaß als Streuungsmaß zu wählen. Eine Möglichkeit besteht etwa darin, die *mittlere absolute Abweichung von dem Median* zu betrachten:

Ermittlung der durchschnittlichen Abweichung aus der Urliste:

$$s_d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}_{50\%}| \quad \bar{x}_{50\%} = \text{Median}$$

Die durchschnittliche Abweichung berücksichtigt sämtliche Beobachtungswerte der Verteilung, indem sie die Abweichungen der Merkmalswerte vom Median in die Berechnung einbezieht. Da diese Abweichungen sowohl positive als auch negative Werte annehmen können, besteht die Gefahr, daß sich positive und negative Abweichungen gegenseitig kompensieren. In der Berechnung wird daher das Vorzeichen der Abweichungen nicht berücksichtigt, d.h. man geht von Absolutbeträgen aus. Der Vorteil dieser Streuungskennzahl besteht darin, daß sie weniger ausreißerempfindlich ist als die im Anschluß zu behandelnde Varianz und die Standardabweichung. Die mittlere Abweichung vom Median wird in der Praxis sehr selten verwendet. Dennoch ist diese Maßzahl zur Messung der Streuung anschaulich zu interpretieren.

Ein Maß für den mittleren Abstand zum arithmetischen Mittel ist die Varianz (Variation; engl.: sum of squares). Da die Summe der Abweichungen der Einzelwerte von ihrem Mittelwert stets den Wert Null annimmt, wird für das Streuungsmaß ein quadratischer Abstand verwendet. Damit werden die Vorzeichen der Abweichungen neutralisiert. Die Varianz ist dann definiert als die mittlere Summe der Abweichungsquadrate

Ermittlung der Varianz aus der Urliste:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \bar{x} = \text{arithmetisches Mittel}$$

Liegt eine Häufigkeitsverteilung vor, dann folgt die Varianz aus dem gewogenen Mittel der Abweichungsquadrate:

Ermittlung der Varianz aus der Häufigkeitstabelle:

$$s^2 = \frac{1}{n} \sum_{j=1}^m h_j \cdot (x_j - \bar{x})^2 \quad \bar{x} = \text{arithmetisches Mittel}$$

Bei klassifizierten Daten mit konstanter Klassenbreite tritt die Klassenmitte an die Stelle der Ausprägung x_j ($j = 1, \dots, m$). Die Varianz ist stets größer oder gleich Null. Nimmt sie den Wert Null an, dann sind sämtliche Merkmalswerte einander gleich, d.h. es liegt eine Konstante vor. Die Quadrierung führt dazu, daß größere Abweichungen erheblich stärker den Wert der Varianz beeinflussen als kleinere Abweichungen. Das bedeutet auch, das Ausreißer ein starkes Gewicht bekommen. Der Unterschied zur mittleren Abweichung vom Median besteht demnach in der Gewichtung der Abweichungen durch die Quadrierung.⁶

Die Varianz kann von ihrer inhaltlichen Interpretation her als Streuungsmaß nicht voll befriedigen. Da die Beobachtungswerte in die Varianzformel quadriert eingehen, unterscheidet sich die Dimension und Größenordnung der Varianz von derjenigen der Beobachtungswerte. Die Varianz ist daher kein anschauliches Maß. Diese Problem wird mit dem Übergang zur Standardabweichung als der Quadratwurzel aus der Varianz behoben:

Ermittlung der Standardabweichung:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{Urliste})$$

\bar{x} = arithmetisches Mittel

Die Standardabweichung ist eine Maßzahl, die die Dimension des erhobenen Merkmals aufweist. Für deskriptive Zwecke ist daher die Standardabweichung der Varianz vorzuziehen, da sie die Streuung in der Maßeinheit der zugrundeliegenden Merkmalsausprägungen ausdrückt.

⁶ Die spezifische Art der Gewichtung durch Quadrieren ist nicht rein willkürlich gewählt. Das Streuungsmaß *Varianz* stammt vielmehr aus der Gleichung der *Normalverteilung*, die in der schließenden Statistik eine überragende Rolle spielt.

Zum Vergleich von Streuungen aus Verteilungen mit unterschiedlichen Dimensionen oder Maßeinheiten sind die Varianz bzw. die Standardabweichung nicht geeignet. Die beiden Streuungskennzahlen hängen in ihrer Größe auch von der Dimension der Merkmalsausprägungen ab. Ist der Mittelwert einer Verteilung z.B. 100 und die Varianz 4,5, so würde man diese als sehr gering bezeichnen. Liegt der Mittelwert einer Verteilung jedoch bei 10, so wäre eine Varianz von 4,5 relativ groß. Diese ungleichen Relationen gelten um so mehr, wenn die Streuung von Merkmalen verglichen werden soll, deren Ausprägungen sich auf sehr unterschiedliche Maßeinheiten beziehen (z.B. wenn die Streuung des Merkmals ALTER in Jahren mit der Streuung des Merkmals EINKOMMEN in DM verglichen werden soll). Für vergleichende Fragestellungen erweist es sich als nützlich, wenn man die Streuung zu dem korrespondierenden Mittelwert in Beziehung setzt. Eine vom Mittelwert bereinigte – und daher dimensionslose – Streuungsmaß ist der *Variationskoeffizient V*:

$$V = \frac{s}{\bar{x}}$$

Der Variationskoeffizient mißt das Verhältnis von Standardabweichung und arithmetischem Mittel. Diese Kennzahl ist nicht mehr in der Maßeinheit des erhobenen Merkmals interpretierbar und damit nicht sehr anschaulich. Seine eigentliche Aussagekraft gewinnt er erst durch den numerischen Vergleich von Variationskoeffizienten aus mehreren unterschiedlichen Verteilungen.

Als empirische Beispiele dienen im folgenden die Sozialstrukturmerkmale des Wahlkreisdatensatzes, die wir bereits bei der Berechnung von Lagemaßen verwendet haben und die NSDAP-Ergebnisse bei den Reichstagswahlen 1930 bis 1933. In der Tabelle 4.3 sind die Werte ausgewählter Streuungsmaße für fünf Strukturmerkmale wiedergegeben. Die höchste Variabilität weisen die Prozentanteile für die Religionszugehörigkeit und die Bevölkerungsanteile in Gemeinden bis zu 5.000 Einwohner auf. Bei dem Merkmal »Prozentanteil Arbeitslose an allen Erwerbspersonen« liegen die Merkmalsausprägungen in der Nähe des Mittelwertes, d.h. die Verteilung ist stark um den Mittelwert zentriert. Insgesamt 50% der Untersuchungseinheiten liegen – bezüglich des Arbeitslosenanteils – zwischen den Quartilswerten 23% (Q_3) und 13% (Q_1). In der Tabelle 4.4 sind die wichtigsten deskriptiven Maßzahlen univariater Verteilungen für die NSDAP-Wahlergebnisse von 1930 bis 1933 zusammengefaßt. Die Werte verdeutlichen für die 35 Wahlkreise den dramatischen Aufstieg der NSDAP. Die Verteilungen für die einzelnen Reichstagswahlen verhalten sich symmetrisch um den Mittelwert, d.h. die Mediane und die arithmetischen Mittel fallen wertemäßig nahezu zusammen. Auffallend ist die relativ geringe Variabilität der NSDAP-Stimmenanteile auf diesem Aggregationsniveau. Bei der Reichstagswahl 1933 weichen z.B. die einzelnen Wahlkreise durchschnittlich um gerundet 7% vom arithmetischen Mittel ab.

Tab. 4.3: Streuungsmaße zu ausgewählten Sozialstrukturmerkmalen auf Wahlkreisebene

Strukturmerkmal	Range	Quartils- abstand	Varianz	Standardab- weichung-
Prozentanteile römisch-katholische Bevölkerung	96 - 3 = 93	48 - 5 = 43	866,5	29,4
Prozentanteile evangelische Bevölkerung	95 - 4 = 91	86 - 49 = 37	750,3	27,4
Prozentanteile Bevölkerung in Gemeinden mit bis zu 5.000 Einwohnern	80 - 0 = 80	58 - 32 = 26	419,7	20,5
Prozentanteile Arbeiter und erwerbslose Arbeiter	52 - 28 = 24	44 - 35 = 9	32,1	5,6
Prozentanteile Arbeitslose an allen Erwerbspersonen	35 - 9 = 26	23 - 13 = 10	36,7	6,1

Tab. 4.4: NSDAP - Wachstum bei den Reichstagswahlen 1930 bis 1933: Deskriptive Maßzahlen der univariaten Verteilungen auf Wahlkreisebene

Wahl- Jahr	Spannweite		Quartile			Standard- Ab- weichung	arith. Mittel
	Min.	Max	25 %	50 %	75 %		
1930	9,4	27,0	14,9	19,2	21,0	4,4	18,5
1932J	20,2	51,0	29,2	39,3	44,8	8,9	37,7
1932N	17,4	45,7	26,2	34,1	40,6	8,0	33,4
1933	30,1	56,5	38,9	45,4	49,4	7,1	44,4

5. Bivariate Verteilungen, Assoziationsmaße und Modelle der proportionalen Reduktion des Vorhersagefehlers

In den vorausgegangenen Abschnitten wurde dargestellt, wie sich die Verteilungen von einzelnen Merkmalen beschreiben lassen und wie die jeweiligen Verteilungsverläufe durch geeignete statistische Maßzahlen in komprimierter Form gekennzeichnet werden können. Derartige Beschreibungen fallen in die Klasse der univariaten Datenanalyse. Sie stellen in der Regel – als 1. Schritt – den Einstieg in die datenanalytische Fragestellung dar, bei der zwei oder mehrere Merkmale im Hinblick auf ihre *gemeinsame Häufigkeitsverteilung* untersucht werden. Für die meisten Problemstellungen folgt der Beschreibung einzelner Merkmale – in einem 2. Schritt – die Diskussion, ob zwischen den betrachteten Merkmalen statistische Zusammenhänge oder Abhängigkeiten be-

stehen und welche Stärke sie gegebenenfalls besitzen. Wir beschränken die folgende Darstellung auf die gemeinsame Verteilung zweier Merkmale. Damit wird der Übergang von der univariaten zur bivariaten Datenanalyse vollzogen. Zur Analyse der Beziehung zwischen zwei Merkmalen eignen sich – je nach Skalenniveau der in die Analyse einbezogenen Merkmale – die Kreuztabellenanalyse, die Varianzanalyse, die Korrelationsanalyse und die Regressionsanalyse. Bei der Analyse zweidimensionaler Verteilungen ergeben sich neue Fragestellungen: Treten die Merkmale unabhängig voneinander auf? Erfolgt das gemeinsame Auftreten der verschiedenen Merkmalsausprägungen systematisch nach bestimmten Mustern? Wie können solche Zusammenhänge bzw. Abhängigkeiten zwischen zwei Merkmalen modellhaft abgebildet werden?

5.1 Zweidimensionale Häufigkeitsverteilungen: Tabellarische und graphische Darstellung

Aufgabe der *bivariaten Statistik* ist die Beschreibung der simultanen Verteilung zweier Merkmale in Form einer *gemeinsamen Häufigkeitsverteilung*. Haben die Merkmale nicht allzuvielen Ausprägungen, dann kann die gemeinsame Verteilung zweier Merkmale in Form einer *Kontingenztafel* dargestellt werden. Sie entsteht durch die Kreuztabellierung zweier Merkmale. Kann inhaltlich zwischen einer abhängigen und einer unabhängigen Variablen unterschieden werden, dann steht üblicherweise im Tabellenkopf die abhängige Variable (in der Statistik allgemein mit dem Großbuchstaben Y symbolisiert) und am linken Tabellenrand die unabhängige Variable (symbolisiert mit dem Großbuchstaben X). Die Variable Y bildet die Spaltenvariable, da jede Merkmalsausprägung y_j ($j = 1, \dots, m$) in einer Spalte der Tabelle angegeben wird (j = Spaltenindex). Die Variable X ist entsprechend die Zeilenvariable, da jede Ausprägung x_i ($i = 1, \dots, k$) in einer Zeile der Tabelle erscheint (i = Zeilenindex). Die Ausprägungen x_i stehen in der Vorspalte der Tabelle. Allgemein ist das Ergebnis der Kreuztabellierung zweier Merkmale X und Y mit den Ausprägungen x_i ($i = 1, \dots, k$) und y_j ($j = 1, \dots, m$) eine Tabelle der Größe $k \times m$, in der die *Tabellenzellen* lokalisiert sind. Die ij -te Zelle ist damit die Zelle der i -ten Zeile und der j -ten Spalte. Die allgemeine Struktur einer bivariaten Häufigkeitstabelle ist in Tabelle 5.1a wiedergegeben.

Die Häufigkeit n_{ij} wird als *Zellenhäufigkeit* bezeichnet. In diesem Zahlenteil der Tabelle stehen die gemeinsamen absoluten Häufigkeiten der paarweisen Kombinationen der Merkmalsausprägungen x_i ($i=1, \dots, k$) und y_j ($j=1, \dots, m$). Die Zellenhäufigkeiten können auch als sogenannte konditionale oder bedingte Häufigkeiten aufgefaßt werden:

- n_{ij} = konditionale Verteilung des Merkmals Y bei gegebener Ausprägung $x = 1$ (»unter der Bedingung $x = 1$ «), $j=1, \dots, m$
- $n_{i\cdot}$ = konditionale Verteilung des Merkmals X bei gegebener Ausprägung $y = 1$ (»unter der Bedingung $y = 1$ «), $i=1, \dots, k$.

Tab. 5.1a: Notation für eine allgemeine bivariate Kontingenztabelle für absolute Häufigkeiten

Y X	y_1	y_2	y_3	\dots	y_j	\dots	y_m	Summe
x_1	n_{11}	n_{12}	n_{13}	\dots	n_{1j}	\dots	n_{1m}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	n_{23}	\dots	n_{2j}	\dots	n_{2m}	$n_{2\cdot}$
x_3	n_{31}	n_{32}	n_{33}	\dots	n_{3j}	\dots	n_{3m}	$n_{3\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_i	n_{i1}	n_{i2}	n_{i3}	\dots	n_{ij}	\dots	n_{im}	$n_{i\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_{k1}	n_{k2}	n_{k3}	\dots	n_{kj}	\dots	n_{km}	$n_{k\cdot}$
Summe	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	\dots	$n_{\cdot j}$	\dots	$n_{\cdot m}$	n

Tab. 5.1b: Die regionale Verteilung der Wahlkreise tabelliert gegen die klassifizierten NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (Angaben absolut)

Großregionen:	Wahlkreise:	NSDAP < / = 45,4 %	NSDAP > 45,4 %	Zeilen- Summen
Norddeutschland	13, 14, 15, 34, 35, 6, 1	2	5	7
Westliche Mitte	16, 17, 18, 19, 20, 21, 22, 23	6	2	8
Östliche Mitte	10, 11, 12, 29, 30, 4, 3, 2, 28, 5, 8, 7, 9	6	7	13
Süddeutschland	33, 27, 32, 31, 26, 25, 24	4	3	7
Spalten-Summen	Gesamtes Reich	18	17	35

Die summierten Häufigkeiten der Zeilen oder Spalten heißen Randhäufigkeiten, marginale Häufigkeiten oder einfach Randsummen und werden entsprechend mit $n_{\cdot j}$ bzw. $n_{i \cdot}$ notiert. Sie entsprechen den univariaten Verteilungen der kreuztabellierten Variablen:

Randverteilung der Variablen X: $n_{i \cdot} = n_{i1} + \dots + n_{im} = n(X=x_i)$, $i = 1, \dots, k$ und

Randverteilung der Variablen Y: $n_{\cdot j} = n_{1j} + \dots + n_{kj} = n(Y=y_j)$, $j = 1, \dots, m$.

Die Zellenhäufigkeiten oder die Häufigkeiten in den beiden Randverteilungen summieren sich jeweils zur Gesamthäufigkeit n (gleich dem Stichprobenumfang) auf:

$$\sum_{i=1}^k \sum_{j=1}^m n_{ij} = \sum_{i=1}^k n_{i.} = \sum_{j=1}^m n_{.j} = n$$

Eine tabellarische Aufbereitung erfordert insbesondere bei metrischen Merkmalen (eventuell auch bei ordinalen Merkmalen mit sehr vielen Ausprägungen) eine Klassenbildung in den Merkmalsausprägungen $i = 1, \dots, k$ bzw. $j = 1, \dots, m$. In diesem Fall stehen in der Kreuztabelle die Klassengrenzen $x_i^U - x_i^O$ bzw. $y_j^U - y_j^O$.

Die Tabelle 5.1b illustriert ein empirisches Beispiel anhand der 35 Wahlkreise der Weimarer Republik, aufgegliedert nach Großregionen (zur regionalen Gruppierung s. Falter, a.a.O., S. 156) und NSDAP-Stimmenanteilen bei der Reichstagswahl 1933. Die Prozentskala der NSDAP-Anteile wurde nach dem Wert des Medians (Stimmenanteil von 45,4%) in zwei Wahlkreisgruppen klassifiziert: 51,4% der 35 Wahlkreise weisen einen Wert kleiner/gleich 45,4% NSDAP-Stimmenanteile auf (insgesamt 18 Wahlkreise), 48,6% der Wahlkreise einen Stimmenanteil von größer 45,4% (insgesamt 17 Wahlkreise). Diese Klassifikation ist selbstverständlich sehr grob, aber aufgrund der sehr geringen Anzahl von Untersuchungseinheiten notwendig. Insgesamt 7 der 35 Wahlkreise lagen in der Weimarer Republik in Norddeutschland. Davon fallen 5 Wahlkreise in das obere Intervall des Merkmals »NSDAP-Stimmenanteile« (d.h. Wahlkreise mit einem Anteil größer 45,4%). In der Großregion »westliche Mitte« kehren sich die Verhältnisse um: der überwiegende Teil der Wahlkreise (6 von insgesamt 8) liegt hier im unteren Intervall (kleiner/gleich 45,4%) der klassifizierten NSDAP-Stimmenanteile. In Süddeutschland und in der östlichen Mitte verteilen sich die Wahlkreise nahezu gleich auf das untere und obere Intervall der klassifizierten NSDAP-Stimmenanteile.

Die tabellarische Aufbereitung erfordert in der Regel auch die Berechnung relativer Häufigkeiten (Anteilswerte bzw. Proportionen). Aus der Division der Zellenhäufigkeiten durch die Anzahl aller Beobachtungswerte (n) ergeben sich die *gemeinsamen relativen Häufigkeiten* r_{ij} :

Zelle(i,j): $r_{ij} = n_{ij} / n$, bzw. in Prozent ausgedrückt: $r_{ij}(\%) = (n_{ij} / n) * 100$.

Die Summe der Zellenprozentwerte ergibt dann 100%. Die Anteilswerte der Randverteilungen erhält man aus der Summenbildung.

Randverteilung der relativen Häufigkeiten von X:

$$r_{i.} = r_{i1} + \dots + r_{im} = n(X = i)/n, i = 1, \dots, k.$$

Randverteilung der relativen Häufigkeiten von Y:

$$r_{.j} = r_{1j} + \dots + r_{kj} = n(Y = j)/n, j = 1, \dots, m.$$

Ein struktureller Zugang zu den in Tabelle 5.1a enthaltenen Informationen ist möglich, indem (a) die Zeilenhäufigkeiten durch die zugehörigen Zeilensummen oder (b) die Spaltenhäufigkeiten durch die jeweiligen Spaltensummen dividiert werden. Hierdurch erhält man sogenannte *bedingte relative Häufigkeitsverteilungen*. Die bedingten relativen Zellenhäufigkeiten erhält man aus den Gleichungen:

- (a) Bedingte relative Häufigkeitsverteilung des Merkmals X bei vorgegebenem y_j :

$$r_{i/j} = n_{ij} / n_j \text{ (für } n_j > 0), i = 1, \dots, k.$$

Bei der bedingten (relativen) Häufigkeit des Merkmals X wird durch die Ergänzung »/y_j« (»unter der Bedingung y_j«) zum Ausdruck gebracht, daß nur Ausprägungen des Merkmals X in Kombination mit einer festen Ausprägung y_j betrachtet werden.

- (b) Bedingte relative Häufigkeitsverteilung des Merkmals Y bei vorgegebenem x_i :

$$r_{j/i} = n_{ij} / n_i \text{ (für } n_i > 0), j = 1, \dots, m.$$

Über eine Multiplikation mit Hundert ergeben sich die bedingten relativen Häufigkeiten in Prozent ausgedrückt:

Spaltenprozente für das Merkmal X:

$$r_{i/j}(\%) = (n_{ij} / n_j) * 100 \text{ (für } n_j > 0), i = 1, \dots, k.$$

Zeilenprozente für das Merkmal Y:

$$r_{j/i}(\%) = (n_{ij} / n_i) * 100 \text{ (für } n_i > 0), j = 1, \dots, m.$$

Für die gemeinsamen relativen Häufigkeiten und die Randverteilungen der relativen Häufigkeiten ergibt sich dann durch Summenbildung folgende Gleichung:

$$\sum_{i=1}^k \sum_{j=1}^m r_{ij} = \sum_{i=1}^k r_{i.} = \sum_{j=1}^m r_{.j} = 1$$

Für die bedingten relativen Häufigkeiten erhalten wir durch Summenbildung die Gleichungen:

$$\begin{aligned} \sum_{i=1}^k r_{i/j} &= 1 & \text{für alle } j &= 1, \dots, m & \text{bzw.} \\ \sum_{j=1}^m r_{j/i} &= 1 & \text{für alle } i &= 1, \dots, k \end{aligned}$$

In tabellarischer Form stellt die Tabelle 5.2a exemplarisch die bedingte Verteilung von Y bei einem gegebenen Wert x_i dar (»unter der Bedingung $X = x_i$ «), d.h. sie enthält spaltenweise für die i-te Ausprägung des Zeilenmerkmals X die bedingten relativen Häufigkeiten für die einzelnen Ausprägungen y_j ($j = 1, \dots, m$) und die entsprechenden Zeilenprozent. In der Tabelle 5.2b sind für die 35 Wahlkreise die bedingten prozentualen Häufigkeiten für die klassifizierten NSDAP-Stimmenanteile aus der Tabelle 5.1b dargestellt. Von den 7 Wahlkreisen in Norddeutschland weisen 71,4% einen NSDAP-Anteil von über 45,4% auf. Von den 8 Wahlkreisen in der westlichen Mitte sind es lediglich 25%, die in die obere Klasse des NSDAP-Anteils fallen. Das Merkmal »Großregion« wird bei dieser Interpretationsweise jeweils konstant gehalten, d.h. unter der Bedingung »Großregion = Norddeutschland« und »Großregion = westliche Mitte« betrachten wir die bedingte prozentuale Verteilung des klassifizierten NSDAP-Merkmals. Die Zeilen-Summen der Randverteilung des Merkmals »Großregion« muß danach zeilenweise den Wert 100% ergeben.

Die Spalten in der Tabelle 5.2c enthalten demgegenüber die prozentuale Verteilung des Merkmals »Großregion« unter der Bedingung der jeweiligen NSDAP-Stimmenanteilsklasse. Die Prozentuierung verläuft hier in umgekehrter Richtung, d.h. spaltenweise. Die Häufigkeiten in den Zellen werden als Prozentanteile an den Randverteilungswerten der zugehörigen Spalte ausgedrückt (hier 18 und 17). Die 17 Wahlkreise, die jeweils einen NSDAP-Anteil größer als 45,4% aufweisen, verteilen sich auf die Großregionen wie folgt: 29,4% dieser Wahlkreise liegen in Norddeutschland, 11,8% in der westlichen Mitte, 41,7% in der östlichen Mitte und 17,6% in Süddeutschland. Mit dieser Vorgehensweise haben wir die zweite Ausprägung des klassifizierten NSDAP-Merkmals konstant gehalten bzw. als Bedingung für die prozentuale Verteilung der Wahlkreise auf die einzelnen Großregionen vorgegeben.

Zweidimensionale Häufigkeitsdarstellungen lassen sich durch Säulendiagramme graphisch darstellen. Die Abbildung 5.1a zeigt eine graphische Darstellung der Daten aus Tabelle 5.1b in Form eines gruppierten Säulendiagramms. Für jede Großregion sind die Säulen für die beiden Größenklassen des gruppierten NSDAP-Stimmenanteils nebeneinander angeordnet. Die Häufigkeiten der Wahlkreise sind an der senkrechten Achse abzulesen. Diese Darstellung erlaubt einen direkten Größenvergleich der gemeinsamen Wahlkreishäufigkeiten für die einzelnen NSDAP-Anteilsklassen in unterschiedlichen Großregionen. Die Abbildung 5.1b illustriert die graphische Übersetzung der Zeilenprozent aus Tabelle 5.2b in Form von gestapelten 100%-Säulendiagrammen. Die Flächenrelationen innerhalb einer Säule, abgebildet durch zwei unterschiedliche Graustufen, spiegeln die unterschiedlichen Prozentanteile jeweils innerhalb einer Großregion wider. Eine Legende informiert über die inhaltliche Bedeutung der Graustufen. Der Vergleich einheitlicher Graustufen zwischen den Großregionen vermittelt einen Eindruck über die prozentualen Größenverhältnisse der NSDAP-Anteilsklassen in ihrer räumlichen Verteilung.

Tab. 5.2a: Notation für eine allgemeine bivariate Kontingenztafel mit bedingten relativen Häufigkeiten des Merkmals Y

X \ Y	Y = 1	...	Y = j	...	Y = m	Summe
absolute Häufigkeiten innerhalb x_i	n_{i1}	...	n_{ij}	...	n_{im}	$n_{i.}$
bedingte Häufigkeiten innerhalb x_i	$n_{i1}/n_{i.}$...	$n_{ij}/n_{i.}$...	$n_{im}/n_{i.}$	1
bedingte prozentuale Häufigkeiten innerhalb x_i	$n_{i1}/n_{i.}$ * 100	...	$n_{ij}/n_{i.}$ * 100	...	$n_{im}/n_{i.}$ * 100	100 %
Kurzschreibweise	$r_{1/i}$...	$r_{j/i}$...	$r_{m/i}$	1

Tab. 5.2b: Die regionale Verteilung der Wahlkreise gegen die klassifizierten NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (relative Häufigkeiten bezüglich des NSDAP-Anteils)

Großregionen:	Wahlkreise:	NSDAP < / = 45,4 %	NSDAP > 45,4 %	Zeilen-Summen
Norddeutschland	13, 14, 15, 34, 35, 6, 1	0,286 28,6 %	0,716 71,4 %	1,00 100 %
Westliche Mitte	16, 17, 18, 19, 20, 21, 22, 23	0,750 75,0 %	0,250 25,0 %	1,00 100 %
Östliche Mitte	10, 11, 12, 29, 30, 4, 3, 2, 28, 5, 8, 7, 9	0,462 46,2 %	0,538 53,8 %	1,00 100 %
Süddeutschland	33, 27, 32, 31, 26, 25, 24	0,571 57,1 %	0,429 42,9 %	1,00 100 %

Tab. 5.2c: Die regionale Verteilung der Wahlkreise gegen die klassifizierten NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (relative Häufigkeiten bezüglich der Großregionen)

Großregionen:	Wahlkreise:	NSDAP < / = 45,4 %	NSDAP > 45,4 %
Norddeutschland	13, 14, 15, 34, 35, 6, 1	0,111 11,1 %	0,294 29,4 %
Westliche Mitte	16, 17, 18, 19, 20, 21, 22, 23	0,333 33,3 %	0,118 11,8 %
Östliche Mitte	10, 11, 12, 29, 30, 4, 3, 2, 28, 5, 8, 7, 9	0,333 33,3 %	0,417 41,7 %
Süddeutschland	33, 27, 32, 31, 26, 25, 24	0,222 22,2 %	0,176 17,6 %
Spalten-Summen	n = 35	1,00 100 %	1,00 100 %

Regionale Verteilung der Wahlkreise bei der Reichstagswahl 1933 Klassifizierte NSDAP-Stimmenanteile

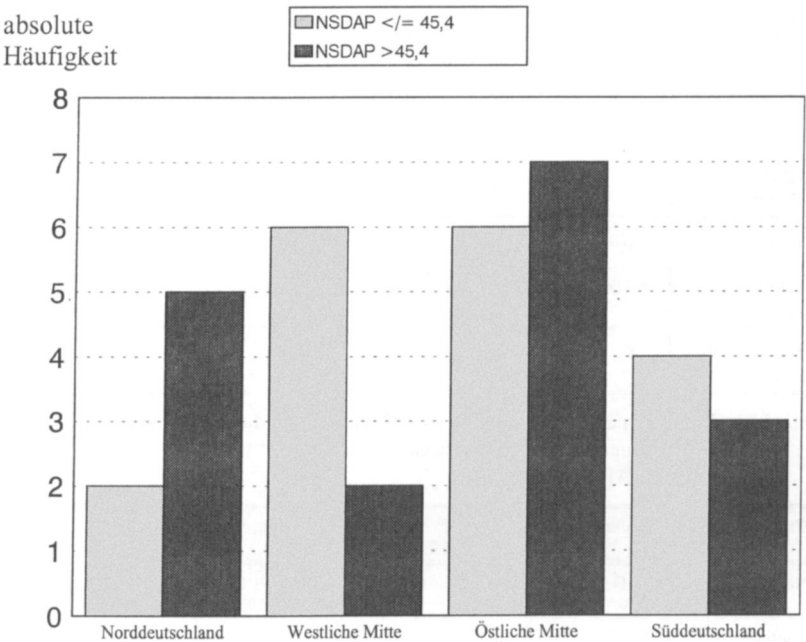


Abb. 5.1a: Gruppiertes Säulendiagramm der Häufigkeitsverteilung zweier Merkmale

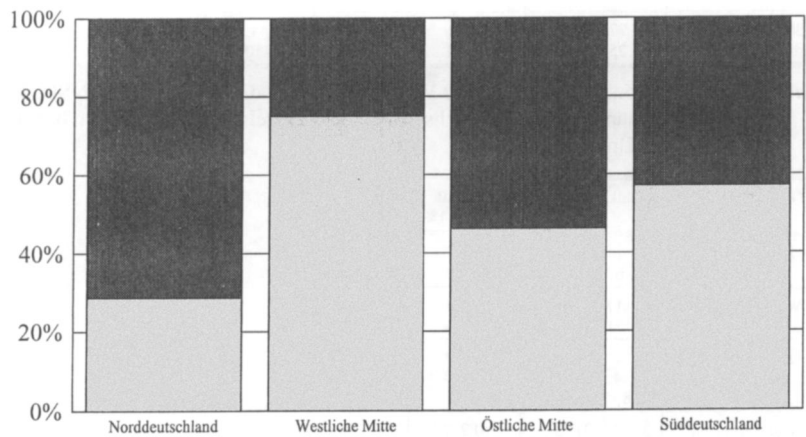


Abb. 5.1b: Gestapeltes Säulendiagramm zur strukturellen Darstellung einer zweidimensionalen Häufigkeitsverteilung

Die eigentliche Analyse von Kontingenztabellen zielt auf die Frage, ob ein Zusammenhang, eine *statistische Beziehung*, zwischen den bivariat tabellierten Merkmalen besteht. Als Oberbegriff für unterschiedliche Modelle der Messung von statistischen Beziehungen wird auch der Ausdruck *Assoziation* verwendet. Das Konzept der statistischen Beziehung beinhaltet die Frage, ob zwei Merkmale gemeinsam oder voneinander unabhängig variieren. Gemeinsame Variation bedeutet z.B. bei zwei ordinalen Merkmalen: Hohe Werte in dem Merkmal X korrespondieren überproportional mit hohen Werten in dem Merkmal Y, niedrige Werte in X korrespondieren überproportional mit niedrigen Werten in Y. In diesem Fall läge eine positive statistische Beziehung vor. Falls die entgegengesetzte Regelmäßigkeit zutrifft, spricht man von einer negativen statistischen Beziehung: Hohe Werte in X korrespondieren überproportional mit niedrigen Werten in Y, niedrige Werte in X korrespondieren überproportional mit hohen Werten in Y. Diese Grundüberlegungen lassen sich anschaulicher auch auf der Grundlage von Prozentanteilen ausdrücken. Die verschiedenen Ausprägungen des unabhängigen Merkmals X sollen hinsichtlich der Verteilung der Werte auf dem abhängigen Merkmal Y verglichen werden. Entsprechend wird die Gesamtzahl der Untersuchungseinheiten für jeden Wert des unabhängigen Merkmals gleich 100% gesetzt. Wird explizit (d.h. theoretisch begründbar) eine Annahme über die Anordnung der beiden Merkmale gemacht und entsprechend zwischen einem unabhängigen und einem abhängigen Merkmal unterschieden, dann gilt für die Kontingenztafel folgende Prozentuierungsregel: Das als unabhängig betrachtete Merkmal wird als Spaltenmerkmal definiert. Ausgehend von der gewählten Merkmalsanordnung in der allgemeinen Kontingenztafel (s. Tabelle 5.1a) bedeutet diese Anordnung für die Prozentuierungsrichtung: Es ist zeilenweise zu Prozentuieren mit den Zeilensummen n_i ($i = 1, \dots, k$) als Prozentuierungsbasis. Entsprechend sind die Prozentwerte der bedingten Verteilungen spaltenweise zu vergleichen. Eine erste visuelle Inspektion dieses Vergleichs kann Anhaltspunkte darüber geben, ob zwischen dem unabhängigen Merkmal X und dem abhängigen Merkmal Y eine statistische Beziehung oder Assoziation besteht.

Allgemein gilt: Zwischen zwei Merkmalen X und Y besteht keine statistische Beziehung, wenn die Prozentsätze für das Merkmal Y gleich sind unter der Bedingung, daß X die Ausprägungen x_j ($j = 1, \dots, k$) aufweist. Zwei Merkmale sind dagegen miteinander assoziiert, wenn die bedingten Verteilungen in Form von Prozentsätzen voneinander abweichen. Mit dieser Aussage ist allerdings noch nichts über die Stärke der beobachteten Assoziation ausgesagt. Mit anderen Worten: Gesucht werden statistische Modelle, die die *Assoziationsstärke* zwischen zwei Merkmalen in Form einer statistischen Maßzahl ausdrücken. Im folgenden wollen wir die wichtigsten Modelle bivariater Assoziation in ihrer wesentlichen Modelllogik und in ihrem jeweiligen Aussageinhalt skizzieren⁷.

⁷ Ausführlich s. Benninghaus, H., 1991: Einführung in die sozialwissenschaftliche Datenanalyse, München/Wien, S. 198ff.

5.2 Die Prozentsatzdifferenz

Einen wichtigen Spezialfall der allgemeinen bivariaten Tabelle (s. Tab. 5.1a) erhält man, wenn die Merkmale X und Y dichotom sind, d.h. nur zwei Ausprägungen aufweisen ($k = 2$, $m = 2$). Prinzipiell kann jedes Merkmal durch Zusammenfassung seiner Ausprägungen auf eine Dichotomie reduziert werden. Ein Beispiel aus der Historischen Wahlforschung soll diese Vorgehensweise illustrieren. Analytische Untersuchungseinheiten sind nun nicht mehr die Wahlkreise der Weimarer Republik, sondern Kreis- und Gemeindeeinheiten in ausgewählten Regionen: Köln-Aachen (65), Oberbayern-Schwaben (89), Niederbayern (56), Franken (92) und Württemberg (120). Die folgenden Auswertungen bewegen sich somit auf der Grundlage von insgesamt 422 Gebietseinheiten. Damit ist eine hinreichend große Datenbasis gegeben, um tiefergehende Unterschiede auf der Ebene der Stadt- und Landkreise oder der Gemeinden für eine differenziertere Analyse von Zusammenhängen bzw. Abhängigkeiten zwischen ausgewählten Strukturmerkmalen und NSDAP-Stimmenanteilen zu berücksichtigen. Zunächst gehen wir lediglich von den Gebieten Köln-Aachen und Franken aus (dichotomes Regionalmerkmal) und von einem sehr grob klassifizierten NSDAP-Stimmenanteil bei den Reichstagswahlen von 1933 (Ausprägung 1: Gebietseinheiten mit einem NSDAP-Anteil kleiner/gleich dem Median von 31,7%; Ausprägung 2: Gebietseinheiten mit einem NSDAP-Anteil größer als der Median). Analog zur generellen Notation für eine $k \times m$ -Kreuztabelle läßt sich auch eine vereinfachende Nomenklatur für eine 2×2 -Tabelle angeben, bei der die absoluten Zellenhäufigkeiten mit den Buchstaben a, b, c und d ausgedrückt werden (s. Tabelle 5.3a). Die Tabelle 5.3b veranschaulicht eine 2×2 -Tabelle, die aus der Kreuztabellierung des dichotomisierten NSDAP-Stimmenanteils mit dem Regionalmerkmal hervorgeht. Das unabhängige Merkmal »Region« steht als Zeilenmerkmal am Tabellenrand, das abhängige Merkmal »NSDAP-Stimmenanteile« als Spaltenmerkmal im Tabellenkopf. Demgemäß informiert die zeilenweise Prozentuierung über die bedingten Verteilungen der NSDAP-Stimmenanteile in den beiden Regionen. In dem Raum Köln-Aachen weisen 87,7% der Gebietseinheiten einen NSDAP-Anteil unterhalb des Medians auf, im fränkischen Raum dagegen sind es lediglich 23,9%. Diese Differenz zwischen den Prozentanteilen ist ein Maß für die Assoziationsstärke zwischen zwei dichotomen Merkmalen. Mit Bezug auf die Nomenklatur der 2×2 -Tabelle kann die Prozentanteilsdifferenz (d%) auch wie folgt berechnet werden:

$$d\% = 100 \left(\frac{a}{a + c} - \frac{b}{b + d} \right)$$

Die Prozentsatzdifferenz vermittelt einen sehr einfachen und anschaulichen Eindruck von der Beziehung zwischen zwei dichotomen (oder dichotomisier-

Tab. 5.3a: Allgemeine Nomenklatur für die Zellenhäufigkeiten einer 2 x 2-Tabelle

Y X	y ₁	y ₂	Summe
x ₁	a	c	a + c
x ₂	b	d	b + d
Summe	a + b	c + d	a+b+c+d (n)

Tab. 5.3b: Ausgewählte Regionen und dichotomisierter NSDAP-Stimmenanteil bei der Reichstagswahl von 1933

Region	NSDAP-Anteil < = 31,7%	NSDAP-Anteil > 31,7%	Zeilen- Summen
Köln-Aachen	57 87,7%	8 12,3%	65 100%
Franken	22 23,9%	70 76,1%	92 100%
Spalten- Summen	79 50,3%	78 49,7%	157 100%

ten) Merkmalen. Die Prozentsatzdifferenz hat bei vollständiger Unabhängigkeit der Merkmale den Wert Null, bei vollständiger Abhängigkeit nimmt sie den Wert + 100 bzw. -100 an. In diesem Fall würden wir nur auf einer Diagonalen Häufigkeiten beobachten, d.h. entweder in den Zellen a und d oder in den Zellen b und c. Die Richtung der Assoziation wird durch das Vorzeichen ausgedrückt: Ein positives Vorzeichen erhält man, wenn die Beziehung entlang der (ad)-Diagonalen verläuft (wie in dem empirischen Beispiel mit $d\% = 87,7\% - 23,9\% = +63,8\%$) ein negatives Vorzeichen folgt aus einem Übergewicht der (bc)-Diagonalen. Die Prozentsatzdifferenz wird häufig auch als Proportionsdifferenz ausgedrückt. In diesem Fall geht man von den bedingten relativen Häufigkeiten aus. Der Wertebereich dieser Maßzahl ist damit zwischen -1 und +1 normiert.

Die Prozentsatzdifferenz ist zwar ein sehr einfaches Modell der Messung eines statistischen Zusammenhangs (und zudem auf 2 x 2-Tabellen beschränkt), dennoch ist diese Maßzahl im Vergleich zu anderen Assoziationsmaßen am anschaulichsten. »Es gibt keinen Grund, die Prozentsatzdifferenz als ein primitives Assoziationsmaß zu betrachten, das eines qualifizierten Forschers unwürdig ist, weil es keinen Grund gibt, ein Konzept seiner klaren Bedeutung wegen abzulehnen. Prozentwerte sind die einzigen Kennwerte, die nicht-professionellen Lesern geläufig sind. Man kann sicher sein, daß jedes andere Assoziationsmaß bei Laien auf größere Verständnisschwierigkeiten

stößt. Da aber viele sozialwissenschaftliche Aussagen und Forschungsberichte an ein nicht einschlägig vorgebildetes Publikum gerichtet sind, sollte man die erhellende Funktion eines leicht verständlichen Assoziationsmaßes nicht unterschätzen. Die Prozentsatzdifferenz ist durchaus geeignet, ein intuitives, wenn nicht fundamentales Verständnis für das Konzept der Assoziation zu vermitteln« (Benninghaus, a.a.O., S. 202).

Bei der Dichotomisierung ursprünglich metrischer Merkmale mit Hilfe des Medians hat die Prozentsatzdifferenz den Vorteil, resistent gegenüber Ausreißern zu sein. Aufgrund der extremen Informationsreduktion liefert diese Maßzahl allerdings in den meisten Fällen kaum ein differenziertes Bild von der Assoziationsstärke zweier metrisch skalierten Merkmale.

5.3 Maßzahlen auf der Basis des CHI-Quadrat-Modells

Das CHI-Quadrat-Modell ist ein Assoziationskonzept, das auf dem *Modell der Abweichung von der statistischen Unabhängigkeit* beruht. Dieser Ansatz vergleicht hypothetische Daten mit den empirisch ermittelten Häufigkeiten einer Kontingenztafel. Als Vergleichsmaßstab dient eine theoretische Kontingenztafel mit Häufigkeiten, die man unter der Annahme statistischer Unabhängigkeit zweier Merkmale erwarten würde. Dabei beruhen die Tabellenwerte auf der Annahme, daß allein die Randverteilungen der Kontingenztafel die empirischen Zellenhäufigkeiten widerspiegeln. In der hypothetischen Tabelle besteht somit keine Assoziation zwischen den Merkmalen X und Y. Die Definition der statistischen Unabhängigkeit zweier Merkmale X und Y sei hier allgemein für eine $k \times m$ -Tabelle wiedergegeben. Zwei Merkmale X und Y heißen statistisch unabhängig, falls

$$n_{ij}^* = n_{i.} * n_{.j} / n, \quad i = 1, \dots, k \text{ und } j = 1, \dots, m$$

gilt. Die Häufigkeit n_{ij}^* ist somit ein theoretischer Wert, der sich in einer zweidimensionalen Häufigkeitstabelle bei statistischer Unabhängigkeit ergeben muß. Es gilt außerdem: sind sämtliche bedingte Verteilungen gleich, so sind sie auch identisch mit der Randverteilung. Will man feststellen, ob eine Kontingenztafel einen Zusammenhang zwischen zwei Merkmalen widerspiegelt, so ist es naheliegend, die empirischen Zellenhäufigkeiten mit denen zu vergleichen, die bei Unabhängigkeit aus den beiden Randverteilungen zu erwarten sind. Ausgehend von der empirischen Häufigkeitstabelle erstellt man somit zunächst eine theoretische Tabelle (auch Indifferenztafel genannt, s. Tabelle 5.4a) unter der Annahme, daß die Merkmale unabhängig voneinander sind. Die Erwartungswerte ($f_{\text{erw.}}$ bzw. e_{ij}) für die einzelnen Zellenhäufigkeiten werden aus dem Produkt der einzelnen Randsummen ermittelt:

$$f_{\text{erw.}} = e_{ij} = (n_{i.} * n_{.j}) / n, \quad i = 1, \dots, k \text{ und } j = 1, \dots, m.$$

Zur Verdeutlichung des CHI²-Modells betrachten wir noch einmal das Regionalmerkmal (jetzt mit 5 Ausprägungen vertreten) und die NSDAP-Stimmenan-

teile (klassifiziert nach Quartilen). Die Tabelle 5.4b zeigt neben den absoluten Häufigkeiten und den Zeilenprozenten auch die aus den Randsummen berechneten theoretischen Werte. Zum Beispiel ergibt sich der theoretische Wert in der ersten Zelle aus $65 \cdot 105 / 422 = 16,2$. Man beachte, daß sich die theoretischen Häufigkeiten zu den Randhäufigkeiten der Tabelle addieren lassen. Aus der Tabelle ist unschwer zu erkennen, daß die beobachteten Zellenhäufigkeiten deutlich von den theoretischen Häufigkeiten abweichen.

Die Abweichungen zwischen den empirischen Häufigkeiten ($f_{emp.}$) und den theoretisch erwarteten Häufigkeiten ($f_{erw.}$ bzw. e_{ij}) bilden die Grundlage für die Größe *CHI-Quadrat* (griech.: χ^2), die auch als quadratische Kontingenz bezeichnet wird:

$$CHI^2 = \chi^2 :$$

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(f_{emp.} - f_{erw.})^2}{f_{erw.}} = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Die quadrierten Differenzen zwischen empirischen und theoretisch erwarteten Häufigkeiten, jeweils dividiert durch die erwartete Häufigkeit, werden über alle Zellen aufsummiert. Der CHI^2 -Wert reflektiert den Grad des Zusammenhangs in einer Kontingenztafel: Je größer der berechnete Wert ist, desto stärker weicht die empirische Tabelle von dem theoretischen Modell der Unabhängigkeit ab. Je höher der Wert, desto größer ist auch der Grad des statistischen Zusammenhangs zwischen den betrachteten Merkmalen. Der Wert Null wird erreicht, wenn eine perfekte statistische Unabhängigkeit zwischen den Merkmalen vorliegt. Aus der Tabelle 5.4b ermitteln wir durch Summenbildung einen CHI^2 von 163,46.

Für den CHI^2 -Wert gibt es allerdings keine Obergrenze. Demzufolge ist eine exakte Interpretation nicht möglich. Der numerische Wert hängt von der Anzahl der Beobachtungswerte n ab. Kreuztabellen mit unterschiedlichen Umfängen n lassen sich daher auch nicht mit Hilfe von CHI^2 vergleichen. Als Maßzahl für die Stärke des Zusammenhangs zwischen zwei Merkmalen ist der CHI^2 -Wert selbst nicht geeignet. Aus diesem Grund wurde eine Vielzahl von Koeffizienten zur *Normierung* von CHI^2 vorgeschlagen. Den einfachsten Weg beschreitet der sog. PHI-Koeffizient, welcher den CHI^2 -Wert in Relation zur Anzahl der Untersuchungseinheiten n setzt:

$$\phi = \sqrt{\frac{CHI^2}{n}}$$

Sind die Merkmale nicht assoziiert, nimmt PHI den Wert 0 an. Für das Beispiel in Tabelle 5.4b ermitteln wir einen Wert von $PHI = 0,622$. Die numerische

Tab. 5.4a: Indifferenztafel der absoluten und erwarteten Häufigkeiten unter der Unabhängigkeitsannahme

Y	y_1	y_2	y_3	...	y_j	...	y_m	Summe
X								
x_1	e_{11}	e_{12}	e_{13}	...	e_{1j}	...	e_{1m}	$n_{1.}$
x_2	e_{21}	e_{22}	e_{23}	...	e_{2j}	...	e_{2m}	$n_{2.}$
x_3	e_{31}	e_{32}	e_{33}	...	e_{3j}	...	e_{3m}	$n_{3.}$
.
.
x_i	e_{i1}	e_{i2}	e_{i3}	...	e_{ij}	...	e_{im}	$n_{i.}$
.
.
x_k	e_{k1}	e_{k2}	e_{k3}	...	e_{kj}	...	e_{km}	$n_{k.}$
Summe	$n_{.1}$	$n_{.2}$	$n_{.3}$...	$n_{.j}$...	$n_{.m}$	n

Tab. 5.4b: Ausgewählte Regionen und NSDAP-Stimmenanteile bei der Reichstagswahl 1933 (Quartile)

Zellen: beobachtete Häufigkeiten Zeilenprozente erwartete Häufigkeiten	NSDAP- Anteil	NSDAP- Anteil	NSDAP- Anteil	NSDAP- Anteil	Zeilen- Summen
	11,4 %- 27,2 %	27,3 %- 33,9 %	34 %- 41,7 %	41,8 %- 79,1 %	
Köln-Aachen	51 78,5 % 16,2	7 10,8 % 16,3	3 4,6 % 16,3	4 6,2 % 16,2	65 100 %
Oberbayern- Schwaben	16 18,0 % 22,1	26 29,2 % 22,4	24 27,0 % 14,1	23 25,8 % 22,1	89 100 %
Niederbayern	12 21,4 % 13,9	23 41,1 % 14,1	17 30,4 % 14,1	4 7,1 % 13,9	56 100 %
Franken	11 12,0 % 22,9	12 13,0 % 23,1	23 25,0 % 23,1	46 50 % 22,9	92 100 %
Württemberg	15 12,5 % 29,9	38 31,7 % 30,1	39 32,5 % 30,1	28 23,3 % 29,9	120 100 %
Spalten-Summen	105 24,9 %	106 25,1 %	106 25,1 %	105 24,9 %	422 100 %

Größe zeigt einen hohen Zusammenhang zwischen der Region und den NSDAP-Stimmenanteilen. Eine exakte Beurteilung dieses Wertes ist allerdings nicht möglich, da allgemein in $k \times m$ -Tabellen ($k, m > 2$) Werte größer Eins auftreten können.

Eine mit der Definitionsformel von PHI algebraisch übereinstimmende, auf die Häufigkeiten einer 2×2 -Tabelle bezogene Formel zur direkten Berechnung der Maßzahl PHI ist der folgende Ausdruck:

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Die Maßzahl PHI erreicht nur im Fall einer 2×2 -Tabelle den Wert 1, wenn CHI^2 seinen maximalen Wert n erreicht. Dies ist dann der Fall, wenn zwei Diagonalzellen der 2×2 -Tabelle unbesetzt sind. Der PHI-Koeffizient sollte nur für 2×2 -Tabellen berechnet werden, da nur in diesem Fall eine auf 1 normierte Obergrenze gegeben ist. Nach der zuletzt angegebenen Formel kann schließlich noch die Richtung des Zusammenhangs ermittelt werden, da nach diesem Ausdruck die Maßzahl PHI einen Wertebereich zwischen -1 und $+1$ aufweist, je nach Übergewicht entlang der (bc) - oder (ad) -Diagonalen.

Die älteste Variante für Tabellen beliebiger Größe ist der von Karl Pearson (1904) vorgeschlagene Kontingenz-Koeffizient C , der wie folgt definiert ist:

$$C = \sqrt{\frac{\text{CHI}^2}{\text{CHI}^2 + n}}$$

Diese Maßzahl kann für Tabellen beliebiger Größe verwendet werden. Sie hat allerdings im Gegensatz zur Maßzahl PHI den Nachteil, niemals den Maximalwert von 1 zu erreichen (d.h. selbst im Falle perfekter Assoziation nicht). Der Maximalwert hängt von der Anzahl der Zeilen und Spalten der jeweiligen Tabelle ab und steigt mit zunehmender Größe der Tabelle. Für eine Vierfeldertabelle kann C z.B. einen maximalen Wert von 0.707 annehmen. Aus diesem Grund ist der Kontingenzkoeffizient zwischen Kreuztabellen mit verschiedener Anzahl von Zellen nicht vergleichbar. Für die Tabelle 5.4b ermitteln wir einen Wert $C = 0,53$. Verdeutlicht man sich den Maximalwert einer 4×4 -Tabelle (0,866) und einer 5×5 -Tabelle (0,894), dann wird ersichtlich, daß der Wert von 0,53 einen relativ starken Zusammenhang zwischen der Region und dem NSDAP-Stimmenanteil aufzeigt.

Eine weitere Variante ist der von Cramer vorgeschlagene Koeffizient V , der auch bei größeren Tabellen immer einen Wert zwischen 0 und 1 annimmt:

$$V = \sqrt{\frac{\text{CHI}^2}{n \min \{k-1, m-1\}}}$$

Der Ausdruck »min« steht für Minimum und besagt, daß zunächst zu prüfen ist, ob die Anzahl der Spalten (m) oder die Anzahl der Zeilen (k) kleiner ist. Der jeweils geringere Wert geht dann in die Berechnung von Cramer's V ein. Da

diese Maßzahl unabhängig ist von der jeweiligen Tabellengröße, ist sie für allgemein $m \times k$ ($m, k > 2$) zu bevorzugen. Für den speziellen Fall einer Vierfeldertabelle ist V mit dem Koeffizienten Φ identisch. Für die Tabelle 5.4b ermitteln wir einen Wert von $V = 0,36$. Für kleinere und mittlere Zeilen- bzw. Spaltenzahlen liefert diese korrigierte Größe einen realistischeren Wert. Auch dieser Wert besagt, daß zwischen der Region und dem NSDAP-Stimmenanteil eine beträchtliche Beziehung besteht. Die Größe der NSDAP-Stimmenanteile weisen mithin eine deutlich unterschiedliche geographische Verteilung auf.

Sämtliche auf χ^2 basierende Maßzahlen sind vorzeichenlose Koeffizienten. Sie sagen somit nichts über die Richtung eines Zusammenhangs aus, was bei nominalen Merkmalen auch nicht möglich ist. Abgesehen von den erläuterten Einschränkungen variieren die Zahlenwerte der Koeffizienten zwischen 0 und 1, wobei jeweils ein Wert von 0 anzeigt, daß keine Assoziation vorliegt.

Das χ^2 -Modell vergleicht die empirischen Häufigkeiten mit den unter der Annahme statistischer Unabhängigkeit erwarteten Häufigkeiten. Die auf χ^2 basierenden Koeffizienten sind allerdings nicht leicht zu interpretieren, da die Stärke der Beziehung als Funktion von χ^2 ausgedrückt wird und nicht auf einem plausiblen Konzept der Assoziation beruht. Die Koeffizienten sind ein reines Rechenkonstrukt. Das folgende Assoziationsmodell der proportionalen Fehlerreduktion hat hingegen eine klare inhaltliche Interpretation und hat über die Tabellenanalyse hinaus Analogien in weiteren statistischen Modellansätzen.

5.4 Modelle der proportionalen Fehlerreduktion

Maßzahlen auf der Grundlage des Modells der proportionalen Fehlerreduktion setzen zunächst voraus, daß in einer bivariaten Tabelle explizit zwischen einem unabhängigen Merkmal X und einem abhängigen Merkmal Y unterschieden werden kann. Das (sogenannte asymmetrische) Modell gibt dann den Grad der Vorhersagbarkeit der abhängigen Variablen auf der Grundlage der Informationen in der unabhängigen Variablen an. Entsprechend dem englischen Terminus *proportional reduction in error* werden diese Koeffizienten auch als PRE-Maße bezeichnet. Diese Maßzahlen beruhen auf einem gemeinsamen Grundgedanken. Sie gehen davon aus, man wolle Ausprägungen des abhängigen Merkmals voraussagen. Ausgangspunkt der Überlegungen ist eine Trefferquote, die man erreicht, wenn man die Häufigkeitsverteilung der abhängigen Variablen kennt. Genauer gesagt geht man von deren Kehrwert aus, d.h. von der Möglichkeit, daß man bei der Vorhersage einen Fehler begeht. Hat man ein weiteres, unabhängiges Merkmal erhoben, kann man aufgrund der Kenntnis der Ausprägungen des unabhängigen Merkmals verbesserte Aussagen treffen, die allerdings immer noch mit Fehlern bestimmter Größe behaftet sind. Jedoch werden die Fehler aufgrund der zusätzlichen Kenntnis des unabhängigen Merkmals geringer ausfallen, und zwar wird sich die Größe des Fehlers bei der Vorhersage der Ausprägungen des abhängigen Merkmals umso mehr verrin-

gern, je stärker die Beziehung zwischen dem unabhängigen und abhängigen Merkmal ist. Die Maße, die sich auf die Logik der proportionalen Reduktion des Fehlers stützen, beruhen darauf, daß sie zwei *Fehlermaße* ins Verhältnis zueinander setzen. Das erste mißt die Größe des Fehlers, der bei einer Vorhersage ohne die zusätzliche Kenntnis des unabhängigen Merkmals auftritt, das andere die Größe des Fehlers, der bei der Vorhersage auftritt, wenn man dabei die Kenntnis des unabhängigen Merkmals nutzt. Die Konstruktion der Koeffizienten beruht auf der allgemeinen Form:

$$\text{PRE-Maßzahl} = ((\text{Irrtum ohne X}) - (\text{Irrtum mit X})) / (\text{Irrtum ohne X})$$

oder abgekürzt mit E (für engl.: Error):

$$\text{PRE-Maßzahl} = (E_1 - E_2) / E_1 .$$

Dabei ist E_1 der minimale Fehler bei der Vorhersage der abhängigen Variablen Y, wenn auf die Information von X verzichtet wird (Regel 1). Der Ausdruck E_2 ist der minimale Fehler, wenn zur Vorhersage von Y die unabhängige Variable X herangezogen wird (Regel 2). Die (relative) Verringerung des Prognosefehlers, die beim Übergang von der 1. zur 2. Prognoseform erzielt werden kann, vermittelt einen Eindruck davon, wie gut man von dem unabhängigen Merkmal auf das abhängige Merkmal schließen kann. Ist Y vollständig durch X bedingt, so ist E_2 gleich 0 und folglich der Wert des PRE-Maßes gleich 1. Sind die Variablen dagegen voneinander unabhängig, so ist E_1 gleich E_2 und das PRE-Maß nimmt den Wert 0 an. Zur Konstruktion eines PRE-Maßes sind offensichtlich drei Elemente erforderlich: Zwei Regeln zur Vorhersage und eine Definition des Vorhersagefehlers.

Speziell für Kontingenztabellen wurden von Goodman und Kruskal ein PRE-Maß vorgeschlagen, das allgemein für $k \times m$ -Tabellen berechnet werden kann und klar zu interpretieren ist. Der Koeffizient wird mit dem griechischen Buchstaben λ (Lambda) bezeichnet. Diese Maßzahl ist speziell auf nominalskalierte Merkmale zugeschnitten, kann aber auch für ordinale Merkmale verwendet werden. Allerdings berücksichtigt diese Maßzahl nicht die zusätzliche Information der Rangordnung, die in ordinalskalierten Daten enthalten ist. Zur Illustration der Grundlogik betrachten wir noch einmal die Tabelle 5.4b. Wollte man für die einzelnen Gebietseinheiten den klassifizierten NSDAP-Stimmenanteil voraussagen und wüßte man lediglich die univariate Verteilung am unteren Tabellenrand, dann treffen wir die sicherste Vorhersage, wenn wir die häufigste Kategorie zur Vorhersage verwenden (hier: das zweite oder dritte Intervall des NSDAP-Anteils). Allerdings werden wir uns dann bei 75% der Gebietseinheiten in der Prognose des NSDAP-Stimmenanteils irren. Wissen wir jedoch die Regionszugehörigkeit, so können wir die Vorhersage verbessern, indem wir bei der Vorhersage des NSDAP-Quartils die bedingten Verteilungen heranziehen: Für Köln-Aachen ist das erste Quartil am häufigsten vertreten, für das fränkische Gebiet das vierte Quartil, usf. Aus diesen Überlegungen läßt

sich die von Goodman und Kruskal entwickelte PRE-Maßzahl *Lambda* ableiten. Im folgenden stellen wir die formale Ableitung für eine allgemeine $m \times k$ -Tabelle dar.

Es ist einleuchtend, daß bei der Vorhersage der Ausprägungen von Y ohne Berücksichtigung der unabhängigen Variablen X der geringste Fehler begangen wird, wenn man zur Vorhersage den Modalwert der Randverteilung von Y heranzieht. Der Vorhersagefehler ist folglich:

Fehler nach Regel 1: $E_1 = n - \max n_j$ (Randverteilung Y).

Regel 1 lautet dann: Für jede Untersuchungseinheit wird die Modalkategorie vorhergesagt. Die Regel zur Vorhersage von Y auf der Basis von X bezieht sich auf die bedingte Verteilung der Untersuchungseinheiten über die Ausprägungen der abhängigen Variablen. Die Y-Werte werden auf der Basis ihrer konditionalen Verteilungen (n_{ij} unter der Bedingung $x_i = 1, \dots, k$) so vorhergesagt, daß der Vorhersagefehler minimal ist. Den geringsten Fehler begeht man bei nominalskalierten Variablen, wenn für jede bedingte Verteilung von Y der bedingte Modalwert herangezogen wird. Der Vorhersagefehler wird hier auf die Randverteilung von X bezogen. Er ergibt sich als Summe der marginalen Häufigkeiten von X abzüglich des jeweiligen Modalwertes der Zeile:

$$E_2 = \sum_{i=1}^k (n_{i.} - \max n_{ij})$$

Regel 2 lautet dann: Für jede Untersuchungseinheit mit der Ausprägung x_i wird die Modalkategorie der zugehörigen konditionalen Verteilung von Y vorhergesagt. Setzt man diese Rechenvorschriften in die allgemeine Formel $(E_1 - E_2)/E_1$ eines PRE-Maßes ein, so erhält man für Lambda:

$$\text{Lambda}_{y.x} = \frac{(n - \max n_j) - \sum_{i=1}^k (n_{i.} - \max n_{ij})}{(n - \max n_j)}$$

wobei n = Gesamthäufigkeit (Stichprobenumfang),
 $\max n_j$ = modale Häufigkeit in der Randverteilung von Y,
 $\max n_{ij}$ = modale Häufigkeit in der i-ten Zeile der Kontingenztafel,
 $n_{i.}$ = Häufigkeit der i-ten Ausprägung von X.

Bei vollständiger Abhängigkeit nimmt Lambda den Wert 1 an, bei vollständiger Unabhängigkeit den Wert 0. Lambda ist ein asymmetrisches Maß: Der Koeffizient ist in der Formel unter der Annahme angegeben worden, daß Y die abhängige Variable ist. Lambda läßt sich allerdings auch in reziproker Weise

berechnen, wenn angenommen wird, daß die abhängige Variable am linken Tabellenrand notiert ist. In diesem Fall wird die Blickrichtung umgekehrt: Y wird als unabhängige und X als abhängige Variable definiert. Diese Variante wird entsprechend als $\text{Lambda}_{x,y}$ bezeichnet:

$$\text{Lambda}_{x,y} = \frac{(n - \max n_i) - \sum_{j=1}^m (n_{ij} - \max n_{ij})}{(n - \max n_i)}$$

wobei n = Gesamthäufigkeit (Stichprobenumfang),

$\max n_i$ = modale Häufigkeit in der Randverteilung von X,

$\max n_{ij}$ = modale Häufigkeit in der j-ten Spalte der Kontingenztafel,

n_{ij} = Häufigkeit der j-ten Ausprägung von Y.

Ein Nachteil von Lambda ist, daß die Vorhersage der Werte des abhängigen Merkmals lediglich auf der Zelle mit dem häufigsten Wert beruht. Bei größeren Tabellen muß daher zwangsläufig ein größerer Vorhersagefehler auftreten, wenn nicht ganz extreme bedingte Verteilungen vorliegen. Läßt sich keine der Variablen als abhängig betrachten, so ist es auch möglich, eine symmetrische Form von Lambda zu berechnen. In diesem Fall werden die Formeln der beiden möglichen asymmetrischen Varianten kombiniert. Auf diese – sehr selten verwendete – Konstruktion wollen wir hier nicht näher eingehen.

Für die Tabelle 5.4b erhalten wir einen Lambda-Wert von 0,13. Die Kenntnis der Regionszugehörigkeit der einzelnen Gebietseinheiten verbessert die Vorhersage der NSDAP-Quartile um 0,13 oder – mit 100 multipliziert – um 13% (gegenüber einer Vorhersage, in der lediglich die Randverteilung der NSDAP-Quartile bei einer Vorhersage berücksichtigt wird). Dieser Wert ist deutlich niedriger als der Wert für Cramer's V (0,36) oder der Wert für den Kontingenzkoeffizienten C (0,53). Die Maßzahl Lambda ist insofern ein eher »konservativer« Koeffizient, da er bei größeren Tabellen tendenziell höhere Vorhersagefehler berücksichtigt. Goodman und Kruskal weisen schließlich selbst darauf hin, daß Lambda den Wert Null annehmen kann, ohne daß eine statistische Unabhängigkeit gegeben ist (d.h. Cramer's V oder C ungleich Null sind). In diesem restriktiven Sinn beantwortet die Maßzahl Lambda die Frage, in welchem Ausmaß die Kenntnis des unabhängigen Merkmals dazu verhilft, das abhängige Merkmal vorherzusagen. Wie es mit der Assoziation in einem anderen statistischen Modellkonzept bestellt ist, ist hier von untergeordneter Bedeutung. Das CHI^2 -Konzept – bzw. die daraus abgeleiteten Maßzahlen – beinhaltet eine völlig andere Modell-Logik.

Die bisher betrachteten Modelle der Assoziation zwischen zwei Merkmalen stützen sich ausnahmslos auf eine zweidimensionale Häufigkeitsverteilung in tabellarischer oder graphischer Aufbereitung. Voraussetzung für eine tabellarische Darstellung metrischer Merkmale (z.B. der NSDAP-Stimmenanteil in

den einzelnen Gebietseinheiten) ist zunächst eine sinnvolle Klassenbildung der Ursprungswerte (hier: Bildung von vier gleich groß besetzten Intervallen, definiert durch die 25%-Quartile). Der Vorteil einer tabellarischen Darstellung von bedingten Prozentverteilungen besteht darin, das sich diese Form der Ergebnisdarstellung auch von einem Nichtstatistiker leicht erschließen läßt. Allerdings ist mit der Klassifikation eines metrischen Merkmals auch ein Informationsverlust verbunden. Die Bildung von Klassenintervallen verdeckt die ursprüngliche Heterogenität der NSDAP-Stimmenanteile auf der Ebene der Stadt- und Landkreise oder der Gemeinden. Wir werden im folgenden das PRE-Modell der *einfachen (oder einfaktoriellen) Varianzanalyse* (kurz: VA) darstellen, in dem der Einfluß eines nominalen/ordinalen (Gruppierungs-) Merkmals auf ein metrisch skaliertes abhängiges Merkmal untersucht wird.

Eine Maßzahl, mit der die Assoziation zwischen einer metrischen abhängigen und einer nominalen unabhängigen Variablen beschrieben werden kann, ist η^2 (griechisch: η^2). Ähnlich anderen Maßzahlen, die dem Modell der proportionalen Fehlerreduktion entsprechen, ist der Wert von η^2 als die Verbesserung der Vorhersagegenauigkeit zu interpretieren. Die Maßzahl η^2 drückt die bei der Vorhersage einer abhängigen metrischen Variablen Y durch Heranziehung einer nominalen unabhängigen Variablen X erzielbare Reduktion des Vorhersagefehlers aus. Allgemein kann die Formel dann wiederum wie folgt ausgedrückt werden:

$$\eta^2 = \frac{E_1 - E_2}{E_1}$$

wobei E_1 den Fehler ohne Gruppierungsmerkmal X ausdrückt und E_2 den Fehler unter Berücksichtigung des Gruppierungsmerkmals. Der Koeffizient kann aus dem Analysemodell der einfachen Varianzanalyse abgeleitet werden. Dieses asymmetrische Modell beantwortet die Frage, ob die unterschiedlichen Ausprägungen x_j ($j = 1, \dots, k$) eines nominalen Merkmals X ein unterschiedliches gruppenspezifisches Mittelwertniveau bezüglich des abhängigen metrischen Merkmals Y bewirken. Bei fehlender Abhängigkeit – so die Modellvorstellung – müßten die Mittelwerte des abhängigen Merkmals in den nach X gebildeten $j = 1, \dots, k$ Teilgruppen identische Werte haben. Eine statistische Beziehung oder Assoziation kann man im Rahmen der Varianzanalyse an den Unterschieden zwischen den Gruppenmittelwerten ablesen. Anstelle einer Kontingenztafel mit Zellohäufigkeiten tritt bei einer einfachen VA eine Mittelwerttafel mit den gruppenspezifischen Mittelwerten des abhängigen Merkmals Y (s. Tabelle 5.5a). Als Beispiel zur Illustration der einfachen Varianzanalyse haben wir in Tabelle 5.5b die durchschnittlichen NSDAP-Stimmenanteile und die Standardabweichungen für die fünf Regionen dargestellt. Eine visuelle Inspektion der Ergebnisse weist den fränkischen Raum als NSDAP-Hochburg aus, während ihre Diaspora im Raum Köln-Aachen lagen.

Tab. 5.5a: Allgemeine Mittelwerttabelle als Grundlage für eine einfache Varianzanalyse

Gruppierungs- merkmal X	x = 1	...	x = j	...	x = k	Gesamt
Gruppenmitt- wert von Y	\bar{y}_1	...	\bar{y}_j	...	\bar{y}_k	\bar{y}
Gruppen- häufigkeit	n_1	...	n_j	...	n_k	n

Tab. 5.5b: Durchschnittliche NSDAP-Stimmenanteile in ausgewählten Regionen bei der Reichstagswahl 1933

Regionen	arithmetisches Mittel	Standard- abweichung	Anzahl der Fälle
Köln - Aachen	24,69	9,13	65
Oberbayern - Schwaben	35,60	8,71	89
Niederbayern	32,29	6,73	56
Franken	43,37	13,79	92
Württemberg	36,36	9,19	120
Gesamt	35,39	10,02	422

Die PRE-Maßzahl für die Größe der Assoziation zwischen dem nominalen/ordinalen Gruppierungsmerkmal und dem metrischen abhängigen Merkmal folgt zum einen aus der Variabilität der Gruppenmittelwerte relativ zum Gesamtmittelwert (Variation zwischen den Gruppen) und zum anderen aus der Variabilität der gruppenspezifischen Y-Werte relativ zu den gruppenspezifischen Mittelwerten (Variation innerhalb der Gruppen). Der Einfluß, den die j-te Ausprägung von X auf Y ausübt, wird als Abweichung des entsprechenden Gruppenmittelwertes vom Gesamtmittelwert definiert und als *erklärte Abweichung* bezeichnet:

Einfluß von x_j (= erklärte Abweichung): $(\bar{y}_j - \bar{y})$.

Die einzelnen Beobachtungswerte y_i ($i = 1, \dots, n$) des abhängigen Merkmals variieren auch innerhalb einer jeden Gruppe x_j ($j = 1, \dots, k$) um den gruppenspezifischen Mittelwert. Diese nicht näher spezifizierbare zusätzliche Variationsquelle wird als Vorhersagefehler definiert und mit *nichterklärter Abweichung* (Fehlervariation) bezeichnet:

Vorhersagefehler (= nichterklärte Abweichung): $(y_{ij} - \bar{y}_j)$.

Der Abstand jedes Beobachtungswertes zum Gesamtmittelwert wird als Gesamtabweichung bezeichnet:

Gesamtabweichung: $(y_{ij} - \bar{y})$.

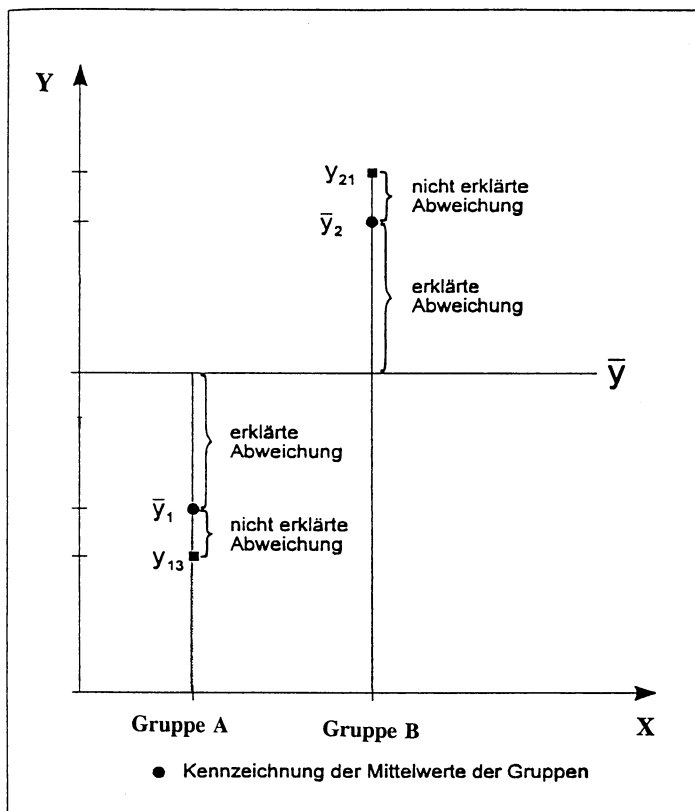


Abb. 5.2: Illustration von erklärter und nicht erklärter Abweichung im Rahmen der einfachen Varianzanalyse

Zur Illustration der erklärten und nicht erklärten Abweichung nehmen wir einmal zwei Ausprägungen des Gruppierungsmerkmals X an und bezeichnen sie mit Gruppe A und Gruppe B. Exemplarisch gehen wir von zwei Untersuchungseinheiten aus: y_{13} bezeichnet den dritten Beobachtungswert in der ersten Gruppe (Gruppe A) und y_{21} bezeichnet den ersten Beobachtungswert in der zweiten Gruppe (Gruppe B). Die zwei Beobachtungswerte erhalten somit zwei Indices: Ein Index gibt die Gruppenzugehörigkeit j an ($j = 1, 2$); der zweite Index zählt die gruppenspezifischen Untersuchungseinheiten (Gruppe A: $i = 3$. Y-Wert, Gruppe B: $i = 1$. Y-Wert). Die Abbildung 5.2 verdeutlicht das Konzept

der erklärten und nichterklärten Abweichungen anhand der beschriebenen Beobachtungswerte y_{13} und y_{21} . Aus den drei Komponenten folgt die vollständige *Zerlegungsgleichung* der VA:

$$\begin{array}{lll} \text{Gesamt-} & = & \text{erklärte} \quad + \text{ nichterklärte} \\ \text{abweichung} & & \text{Abweichung} \quad \text{Abweichung} \\ (y_{ij} - \bar{y}) & = & (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j) . \end{array}$$

Diese Abstände werden jeweils quadriert und als Abstandsquadrate über sämtliche Untersuchungseinheiten aufsummiert. Als Ergebnis erhalten wir die Gleichung der Streuungszерlegung für eine einfache Varianzanalyse:

$$\begin{array}{lll} \text{Gesamtvariation} & = & \text{erklärte Variation} \quad + \text{ nicht erklärte Variation} \\ \text{Summe der quadrierten} & = & \text{Summe der quadrierten} \quad + \text{ Summe der quadrierten} \\ \text{Abweichungen der} & & \text{Abweichungen zwischen} \quad \text{Abweichungen innerhalb} \\ \text{Y-Werte vom Gesamt-} & & \text{den Gruppen} \quad \text{der Gruppen} \\ \text{durchschnitt} & & \end{array}$$

Die Grundgleichung der Variationszerlegung lautet dann formal:

$$\begin{array}{l} \text{Gesamtvariation} = \text{Erklärte Variation} + \text{Nicht erklärte Variation} \\ \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 \end{array}$$

Wenn beide Seiten dieser Gleichung durch die Gesamtvariation dividiert werden, erhält man den *erklärten* und den *nichterklärten Anteil* der Variation von Y:

$$\frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\bar{y}_j - \bar{y})^2}{\sum (y_i - \bar{y})^2} + \frac{\sum (y_i - \bar{y}_j)^2}{\sum (y_i - \bar{y})^2}$$

$$\frac{\text{Gesamtvariation}}{\text{Gesamtvariation}} = \frac{\text{Erklärte Variation}}{\text{Gesamtvariation}} + \frac{\text{Nicht erklärte Variation}}{\text{Gesamtvariation}}$$

$$1 = \text{Variationsanteil,} + \text{Variationsanteil,}$$

$$\qquad \qquad \text{der erklärt ist} \qquad \qquad \text{der nicht erklärt ist}$$

$$1 = \eta^2 + 1 - \eta^2$$

Die auf die Maßzahl ETA^2 zugeschnittenen Vorhersageregeln und Fehlerdefinitionen des PRE-Modells lassen sich nun wie folgt angeben:

Erste Vorhersageregeln: Vorhersage der abhängigen Variablen auf der Grundlage ihrer eigenen Verteilung. Das Gesamtmittel wird zur Vorhersage der einzelnen Beobachtungswerte y_i ($i = 1, \dots, n$) herangezogen.

Erste Fehlerdefinition E_1 : Summe der quadrierten Abweichungen der Y-Werte vom Gesamtmittelwert.

Zweite Vorhersageregeln: Vorhersage der abhängigen Variablen auf der Grundlage der gruppenspezifischen Mittelwerte. Zur Vorhersage der gruppenspezifischen Beobachtungswerte y_{ij} ($j = 1, \dots, n_j$) werden die Gruppenmittelwerte verwendet.

Zweite Fehlerdefinition E_2 : Summe der quadrierten Abweichungen der gruppenspezifischen Y-Werte vom jeweiligen gruppenspezifischen Mittelwert.

Der generelle Ausdruck zur Berechnung der proportionalen Fehlerreduktion lautet dann für die einfache Varianzanalyse:

$$\eta^2 = \frac{E_1 - E_2}{E_1} = \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \bar{y}_j)^2}{\sum (y_i - \bar{y})^2}$$

$$= \frac{(\text{Gesamtvariation} - \text{Nicht erklärte Variation})}{\text{Gesamtvariation}}$$

Die Maßzahl η^2 kann Werte zwischen 0 und 1 annehmen. Mit 100 multipliziert gibt sie den Prozentanteil der Varianz von Y an, der durch die Einführung des Gruppierungsmerkmals X erklärt wird. Ein η^2 -Wert von Null bedeutet gleiche Mittelwerte in den einzelnen Gruppen, ein η^2 -Wert von Eins bedeutet, daß nach Einführung des Gruppierungsmerkmals keinerlei Varianz innerhalb der Gruppen zu beobachten ist. Mit anderen Worten: In diesem Fall erklärt das Gruppierungsmerkmal die Y-Varianz vollständig im Sinne einer perfekten Vorhersage.

Verwenden wir das PRE-Modell der Streuungszerlegung zur Feststellung des Ausmaßes regionaler Unterschiede in den NSDAP-Stimmanteilen der 422 Stadt-/Landkreise und Gemeinden (s. die Ausgangstabelle 5.5b), so ergibt sich folgender empirische Befund (s. Tabelle 5.6): Das Regionalmerkmal erklärt 25% der ursprünglichen Varianz in der Verteilung des metrischen Merkmals »NSDAP-Stimmenanteile von 1933«. Das unabhängige Regionsmerkmal besitzt demnach eine relativ hohe Erklärungskraft. Zum Vergleich: Die bivariate tabellarische Auswertung des klassifizierten NSDAP-Stimmenanteils (1933) ergab gegen das Regionsmerkmal kreuztabelliert in einer 5×4 -Tabelle einen Lambda-Wert von 0,13, d.h. die proportionale Fehlerreduktion ist hier mit 13% deutlich niedriger als im Rahmen des PRE-Modells der Streuungszerlegung! Die Fehlerquote ist modellimmanent bei der Maßzahl Lambda im Fall größerer

Tab. 5.6: Variationszerlegungstabelle für die regionale Verteilung der NSDAP-Stimmenanteile bei der Reichstagswahl 1933

Variations- quelle	Summe der Abweichungs- quadrate	ETA ²
Zwischen den Gruppen	13958,43	13958,43/55822,31 = 0,25
Innerhalb der Gruppen	41863,88	
Gesamt	55822,31	---

Kreuztabellen tendenziell hoch, da lediglich ein Wert – jeweils der Modalwert aus den bedingten Verteilungen – zur Vorhersage berücksichtigt wird. Demgegenüber wird in dem PRE-Modell der einfachen Varianzanalyse der vollständige Skalenumfang des metrischen abhängigen Merkmals ausgeschöpft. Da die Variationen sowohl auf der Ebene der Gruppenmittelwerte als auch auf der Ebene der einzelnen Untersuchungseinheiten (d.h. hier der Gebietseinheiten) berücksichtigt werden, werden in dem PRE-Modell der VA wesentlich mehr Informationen bei der Berechnung der Maßzahl berücksichtigt als bei einem klassifizierten Merkmal in Form der Modalwerte. Bei einem metrischen abhängigen Merkmal sollte man daher eine Klassifikation nur für eine tabellarisch übersichtliche, verdichtete Darstellung vorsehen. Als Maßzahl ist in diesem Fall ETA² für die ursprünglichen Ausprägungen des abhängigen Merkmals das angemessenere PRE-Modell.

5.5 Die Analyse der Beziehung zwischen ordinalen Merkmalen

Die bisher diskutierten Koeffizienten sind primär Maßzahlen für die Stärke eines Zusammenhangs (bzw. einer Abhängigkeit) für nominales Datenniveau. Eine Ausnahme bildet das PRE-Modell der einfachen Varianzanalyse, das für das abhängige Merkmal metrisches Skalenniveau voraussetzt. Handelt es sich bei den zu analysierenden Merkmalen um solche mit ordinalen Meßniveaus, würde man bei Rückgriff auf diese Maßzahlen Informationen verschenken. Zusätzlich zur Information über einen Unterschied von Ausprägungen kann man ordinalskalierte Daten in eine eindeutige Rangfolge ordnen. Im Gegensatz zu Kreuztabellen mit nominalen Merkmalen kann man im Falle ordinaler Merkmale Zusammenhangsmaße bilden, die auch Auskunft über die Richtung des Zusammenhangs geben, angezeigt durch das Vorzeichen dieser Assoziationsmaße. Zwei ordinale Merkmale sind positiv assoziiert, wenn niedrige Werte auf dem einen Merkmal auch niedrige auf dem anderen nach sich ziehen und

hohe Werte auf dem ersten Merkmal auch hohe auf dem zweiten Merkmal. Umgekehrt sind zwei ordinale Merkmale negativ assoziiert, wenn niedrige Werte auf dem einen Merkmal tendenziell mit hohen Werten auf dem anderen verbunden sind. Maßzahlen für ordinale Merkmale verdeutlichen die Richtung des Zusammenhangs, indem sie einen Wertebereich zwischen -1 (negative Assoziation) und $+1$ (positive Assoziation) annehmen. Der Wert 0 besagt, wie auch bei den bislang dargestellten Maßzahlen, daß kein Zusammenhang zwischen den Merkmalen besteht.

Die im folgenden darzustellenden ordinalen Assoziationsmaße beruhen auf dem Vergleich der Ausprägungen zweier Merkmale für alle möglichen Paare zwischen den Untersuchungseinheiten (*Modell des Paarvergleichs*). Dabei wird bei jedem Paar festgestellt, in welcher Beziehung die Werte zueinander stehen. Von wesentlicher Bedeutung sind hier die Begriffe *konkordant* und *diskonkordant*. Sind z.B. beide Werte der ersten Untersuchungseinheit (x_1, y_1) höher als die Werte der zweiten Untersuchungseinheit (x_2, y_2) oder sind sie umgekehrt beide niedriger, dann spricht man von einem konkordanten Paar (gleichgerichteter Zusammenhang). Beispiel: Einheit (x_1, y_1) weist bezüglich der Variablen X und Y den Wert 4 auf; Einheit (x_2, y_2) weist in den Variablen X und Y den Wert 2 auf. Dieses Beispiel entspricht einer positiven (konsistenten) Beziehung zwischen X und Y. Das Ergebnis ist diskonkordant (entgegengerichteter Zusammenhang), wenn die erste Untersuchungseinheit bei einem Merkmal eine höhere Ausprägung aufweist und bei der anderen Variablen einen niedrigeren Wert hat als die zweite Untersuchungseinheit. Beispiel: Einheit (x_3, y_3) weist bezüglich des Merkmals X den Wert 1 auf und bezüglich des Merkmals Y den Wert 3; Einheit (x_4, y_4) weist dagegen in beiden Variablen den Wert 2 auf. Dieses Beispiel entspricht einer negativen (inversen) Beziehung zwischen X und Y.

Wenn wir den Vergleich zweier Untersuchungseinheiten auf zwei ordinale Merkmale beziehen, kann man ein Paar von Untersuchungseinheiten wie folgt betrachten:

Wir setzen zunächst voraus, daß sowohl die Rangzahlen der X-Werte als auch die der Y-Werte jeweils verschieden sind (*Rangzahlen ohne Bindung*). Zunächst müssen die n Beobachtungspaare $(x_i, y_i; i = 1, \dots, n)$ nach steigenden X-Werten, also nach den Rangzahlen der X-Werte geordnet werden. Wir gehen ferner davon aus, daß die Beobachtungsreihe bereits in dieser Reihenfolge dargestellt ist. Die zweidimensionale Reihe

$$(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_n, y_n)$$

sei also bereits nach aufsteigenden X-Werten (wachsenden Rangzahlen R der X-Werte) geordnet mit

$$x_1 < x_2 < \dots < x_n, \text{ d.h.: } R(x_1) < R(x_2) < \dots < R(x_n).$$

Im allgemeinen sind darin die Y-Werte nicht der Größe nach geordnet. $R(y_i)$ sei der Rang von y_i für $i = 1, \dots, n$. Ein fiktives Zahlenbeispiel mit $n = 5$ soll diese

Tab. 5.7a: Fiktives Zahlenbeispiel zur Illustration von konkordanten und diskordanten Paaren

UE i	X- Werte	Y- Werte
1	1	3
2	2	2
3	3	1
4	4	4
5	5	5

Tab. 5.7b: Ergebnisse der Paarvergleiche (fiktive Daten)

Bezeichnung	Beschreibung	Anzahl der Untersuchungs- einheiten	Paare
Konkordant falls	$x_i < x_j$ und $y_i < y_j$ oder $x_i > x_j$ und $y_i > y_j$	7	1,2; 1,3; 1,4; 1,5; 2,3; 2,4; 2,5
Diskonkordant falls	$x_i < x_j$ und $y_i > y_j$ oder $x_i > x_j$ und $y_i < y_j$	3	3,4; 3,5; 4,5

Vorgehensweise verdeutlichen (s. Tabelle 5.7a). In der Folge der Rangzahlen von Y wird dann für jede Rangzahl $R(y_i)$, $i = 1, \dots, n$, festgestellt, wie viele der nachfolgenden Rangzahlen kleiner bzw. größer sind. Die Anzahlen werden mit q_1, \dots, q_n bzw. p_1, \dots, p_n bezeichnet, wobei p_i für ein konkordantes Paar steht und q_i für ein diskordantes Paar. Insgesamt ergeben sich $n(n-1)/2$ Paarvergleiche. Die Summen P und Q sind gegeben mit:

$$Q = \sum_{i=1}^n q_i, \quad P = \sum_{i=1}^n p_i.$$

Unter der Voraussetzung von Rangzahlen ohne Bindung sind sämtliche Ränge der Y-Werte verschieden. Somit werden in der Gesamtsumme $P + Q$ sämtliche $n(n-1)/2$ möglichen Paare von Y-Werten genau einmal gezählt. Dann gilt:

$$P + Q = \frac{n(n-1)}{2}.$$

Für das kleine Zahlenbeispiel ergeben sich danach 10 Paare, die danach untersucht werden, ob es jeweils ein konkordantes oder diskordantes Paar ist. Die Ergebnisse des paarweisen Vergleichs finden sich in Tabelle 5.7b. Überwiegen die konkordanten Paare, so ist der Zusammenhang zwischen den Variablen positiv; ein negativer Zusammenhang ergibt sich aus dem Überwiegen der diskordanten Paare. Im Zähler enthalten die Maßzahlen ordinaler Assoziation daher den Ausdruck $(P - Q)$. Die Maßzahlen unterscheiden sich in der Normierung dieser Differenz: den Rangkorrelationskoeffizienten von Kendall (Kendalls Tau_a ; griech.: τ_a) bei Beobachtungsreihen ohne Bindungen wird der Ausdruck durch $(P + Q)$ dividiert:

$$\text{Tau}_a = \frac{P - Q}{P + Q} = \frac{P - Q}{\frac{1}{2} \cdot n(n - 1)} = \frac{2(P - Q)}{n(n - 1)}$$

Ein Koeffizient von $\text{Tau}_a = 1$ ist genau dann erfüllt, wenn $Q = 1$ ist. Dann müssen alle Q_i ($i = 1, \dots, n$) verschwinden, d.h. alle Y-Werte mit einem kleineren Rang als $R(y_i)$ müssen auch links von y_i liegen. Somit sind die Rangzahlen der Y-Werte genauso wie die der X-Werte der Reihe nach steigend. Nur bei völlig gleichgerichteten Rangzahlen ist Tau_a gleich Eins. Ein Koeffizient von $\text{Tau}_a = -1$ ist genau dann erfüllt, wenn P gleich Null ist. Dann sind die Ränge der Y-Werte im Gegensatz zu denen der X-Werte der Reihe nach fallend. Die Ränge verhalten sich dann völlig gegenläufig. Der Wert $\text{Tau}_a = 0$ ist gleichbedeutend mit $P = Q$. In diesem Fall ist zwischen den Rängen kein Zusammenhang erkennbar. Für das fiktive Zahlenbeispiel erhält man z.B. ein Tau_a von 0.40. Der Koeffizient Tau_a nimmt nur dann Werte zwischen -1 und $+1$ an, wenn sämtliche Vergleiche in konkordante bzw. diskordante Paare unterschieden werden können. Dies ist nicht immer der Fall. Weist ein Paar von Untersuchungseinheiten bei dem Vergleich auf einem der beiden Merkmale X oder Y oder auf beiden den gleichen Wert auf, so ist dieses Paar nicht eindeutig zu klassifizieren. Man spricht in diesem Zusammenhang von Bindungen, Verknüpfungen oder *Ties*. Die Tabelle 5.8a zeigt ein Beispiel für ein Paar (x_3, x_5) das bezüglich der X-Variablen verknüpft ist und bezüglich des Y-Merkmals nicht verknüpft ist. Die Tabelle 5.8b zeigt allgemein die Ergebnisse aus Paarvergleichen in Kontingenztabellen unter der Berücksichtigung von Ties.

Treten Ties auf, so ist der Wertebereich des Koeffizienten Tau_a eingeschränkt. Damit ist ein konkreter Wert für Tau_a nicht mehr eindeutig interpretierbar. Für den Fall von Ties sind andere Maßzahlen vorgeschlagen worden, die bei der Normierung der Differenz $(P - Q)$ diese Bindungen auf unterschiedliche Weise explizit berücksichtigen. Es lassen sich folgende Verknüpfungstypen unterscheiden:

Tab. 5.8a: Beispiel für eine Verknüpfung des X-Merkmals

x_i	X- Werte	Y- Werte
3	6	3
5	6	5

Tab. 5.8b: Mögliche Ergebnisse von Paarvergleichen in Kontingenztabelle

Bezeichnung	Beschreibung	Anzahl der Untersuchungs- einheiten
Konkordant falls	$x_i < x_j$ und $y_i < y_j$ oder $x_i > x_j$ und $y_i > y_j$	P
Diskonkordant falls	$x_i < x_j$ und $y_i > y_j$ oder $x_i > x_j$ und $y_i < y_j$	Q
Eine Bindung in dem Merkmal X	$x_i = x_j$ und $y_i < y_j$ oder $y_i > y_j$	T_x
Eine Bindung in dem Merkmal Y	$y_i = y_j$ und $x_i < x_j$ oder $x_i > x_j$	T_y
Eine Bindung in beiden Merkmalen	$x_i = x_j$ und $y_i = y_j$	T_{xy}

- Verknüpft in X (T_x): Untersuchungseinheiten, die im Hinblick auf Y verschieden sind, jedoch im Hinblick auf X verknüpft sind.
- Verknüpft in Y (T_y): Die Untersuchungseinheiten sind im Hinblick auf X verschieden, allerdings im Hinblick auf Y verknüpft.
- Verknüpft in X und Y (T_{xy}): Die Untersuchungseinheiten können im Hinblick auf X und Y verknüpft sein.

Im folgenden wird die Berechnung der konkordanten bzw. diskonkordanten Paare sowie der Ties in einer Kontingenztabelle kurz erläutert. Zur Berechnung der konkordanten Paare multipliziert man jede Zelle mit denjenigen Zellen, die rechts unterhalb von ihr liegen und summiert diese Produkte auf. Zur Berechnung der diskonkordanten Paare verfährt man umgekehrt und multipliziert jede

Zelle mit denjenigen, die links unterhalb von ihr liegen. Zur Berechnung der Ties von X bzw. Y werden sämtliche Zellenhäufigkeiten berücksichtigt. Dabei wird die Zahl der Beobachtungen einer Zelle mit der jeweils benachbarten Zellenhäufigkeit in einer Zeile bzw. Spalte multipliziert und die Ergebnisse addiert.

Am Beispiel einer allgemeinen 2×3 -Tabelle (s. Tabelle 5.9a) veranschaulicht die Tabelle 5.9b die Bildung der insgesamt fünf möglichen Paartypen. Die Buchstaben a bis f bezeichnen die Zellenhäufigkeiten. Das Beispiel in Tabelle

Tab. 5.9a: Allgemeine 2×3 - Tabelle

Y-Variable	x = 1	x = 2	x = 3
y = 1	a	b	c
y = 2	d	e	f

Tab. 5.9b: Ermittlung der Paare in einer allgemeinen 2×3 - Tabelle

Paartyp	Symbol	Anzahl der Paare
konkordant	P	$a(e + f) + b(f)$
diskonkordant	Q	$c(d + e) + b(d)$
nur in X verknüpft	T_x	$a(d) + b(e) + c(f)$
nur in Y verknüpft	T_y	$a(b + c) + b(c)$ $+ d(e + f) + e(f)$
in X und Y verknüpft	T_{xy}	$1/2[a(a-1)$ $+ b(b-1)$ $+ c(c-1)$ $+ d(d-1)$ $+ e(e-1)$ $+ f(f-1)]$
mögliche Paare insgesamt	$P + Q + T_x + T_y + T_{xy}$	$n(n-1)/2$

Quelle: Benninghaus, H., 1991: Einführung in die sozialwissenschaftliche Datenanalyse, 2. überarb. u. erw. A., München, S. 242f.

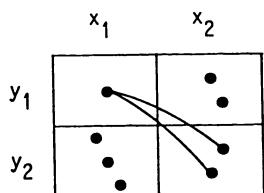
5.10 zeigt die Verteilung der Paare in einer 2×2 Tabelle mit fiktiven Zellenhäufigkeiten (Notation: N_c = konkordante Paare, N_d = diskonkordante Paare).

Die nachfolgenden ordinalen Assoziationsmaße haben verschiedene Nennerausdrücke, in denen die unterschiedliche Behandlung der drei Paar-Typen mit Ties zum Ausdruck kommen. So berücksichtigt Tau_b die Verknüpfungen, falls sie für X und Y auftreten, nicht aber Ties auf beiden Merkmalen X und Y (T_{xy}):

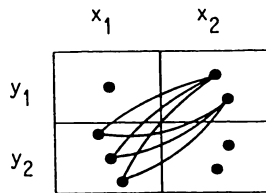
Tab. 5.10: Die Verteilung der Paare in einer 2 x 2 - Tabelle mit fiktiven Zellenbesetzungen

	x_1	x_2	
y_1	1	2	3
y_2	3	2	5
	4	4	8

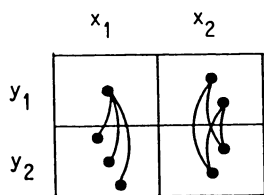
$$\frac{N(N-1)}{2} = \frac{8(8-1)}{2} = 28$$



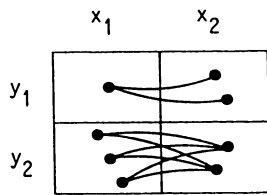
$$N_c = (1)(2) = 2$$



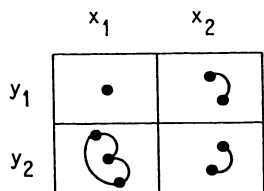
$$N_d = (2)(3) = 6$$



$$T_x = (1)(3) + (2)(2) = 7$$



$$T_y = (1)(2) + (3)(2) = 8$$



Die Anzahl der Paare in Zelle₂₁ ist $\frac{n(n-1)}{2} = \frac{3(3-1)}{2} = 3$ und in Zelle₁₂ und Zelle₂₂ je $\frac{2(2-1)}{2} = 1$.
Folglich ist $T_{xy} = 1 + 3 + 1 = 5$.

$$\binom{N}{2} = \frac{N(N-1)}{2} = N_c + N_d + T_x + T_y + T_{xy} = 2 + 6 + 7 + 8 + 5 = 28$$

Quelle: Benninghaus, H., 1991: Einführung in die sozialwissenschaftliche Datenanalyse, 2. überarb. u. erw. A., München, S. 244f.

$$\text{Tau}_b = \frac{(P - Q)}{\sqrt{(P + Q + T_x) (P + Q + T_y)}}$$

Bei quadratischen Tabellen kann Tau_b die maximalen Werte von +1 und -1 erreichen. Im Hinblick auf das Problem nicht-symmetrischer Kontingenztafeln hat Kendall einen dritten Koeffizienten Tau_c vorgeschlagen, der bei der Normierung die Zahl der Spalten bzw. Zeilen der Tabelle berücksichtigt:

$$\text{Tau}_c = \frac{2 m (P - Q)}{n^2 (m - 1)}$$

Der Buchstabe m steht für die kleinere Anzahl der Zeilen oder Spalten einer Kontingenztafel. Diese Maßzahl kann für beliebige Tabellenformen berechnet werden, ohne daß der Wertebereich der Maßzahl eingeschränkt ist. Der Koeffizient kann die Extremwerte -1 und +1 auch in nicht-quadratischen Tabellen erreichen. Bei vielen Tabellen nimmt Tau_c höhere Werte als Tau_b an. In der sozialwissenschaftlichen Tabellenanalyse ist die Verwendung von Tau_c allerdings verbreiteter.

Eng mit den diskutierten Koeffizienten verwandt ist das von Goodman und Kruskal vorgeschlagene und häufig verwendete *Gamma* (griech.: γ):

$$\text{Gamma} = \frac{P - Q}{P + Q}$$

Gamma ist ein symmetrisches Maß, das für Tabellen beliebiger Größe berechnet werden kann und zwischen +1 und -1 variiert. Gamma nimmt den Wert 1 an, wenn sämtliche Untersuchungseinheiten in den Zellen der Diagonalen einer Tabelle liegen. Wie ersichtlich, berücksichtigt dieser Koeffizient keine Ties. Der Koeffizient nimmt daher stets die höchsten Werte von den ordinalen Maßzahlen an. Gamma ist nicht nur aus diesem Grund ein beliebter Koeffizient, sondern auch aufgrund seiner PRE-Interpretation.

Der Gamma verwandte Koeffizient *Somer's d* ist im Gegensatz zu diesem ein asymmetrisches Maß und setzt daher zunächst die Festlegung des als abhängig zu betrachtenden Merkmals voraus. Im Nenner steht daher die Anzahl aller Paare, die nicht auf dem unabhängigen Merkmal gebunden sind. Ties werden nur für das abhängige Merkmal korrigiert.

$$d_y = \frac{P - Q}{P + Q + T_y}$$

Somer's d gibt somit den Anteil an, um den die konkordanten die diskordanten Paare übersteigen, bezogen auf sämtliche Paare, die nicht auf X gebunden sind. Ein Nachteil von Somer's d besteht darin, daß dieser Koeffizient nur bei quadratischen Tabellen Werte zwischen +1 und -1 annehmen kann. Die Maßzahl Somer's d steht in folgender Beziehung zum Kendallschen Tau_b:

$$\text{Tau}_b = \pm \sqrt{d_{yx} \cdot d_{xy}}$$

Außerdem ist d_y in 2×2 -Tabellen gleich der Prozentsatzdifferenz, d.h. es besteht die Identität $d\% = 100 \cdot d_y$:

$$100 \left(\frac{a}{a + c} - \frac{b}{b + d} \right) = 100 \frac{ad - bc}{(a + c)(b + d)} = 100 \frac{P - Q}{P + Q + T_y}$$

Bei der Tabelle, die im folgenden als Beispiel dienen soll, handelt es sich um eine Kreuztabellierung des Merkmals »NSDAP-Stimmenanteil« auf Gemeinde-/Kreisebene (wiederum klassifiziert nach den Quartilen) und der Konfessionsstruktur (Anteil der römisch-katholischen Bevölkerung). Der Katholikenanteil wurde zunächst in drei Gruppen klassifiziert: überwiegend evangelisch (unter 34% der Bevölkerung katholisch), konfessionell gemischt (zwischen 34% und unter 67% der Bevölkerung katholisch) und überwiegend katholisch (67% und mehr der Bevölkerung katholisch). Diese Reduzierung der metrischen Originalmerkmale auf diskrete ordinale Neuvariablen ermöglicht eine anschauliche, gut mittelbare Kreuztabulation (s. Tabelle 5.11).

Ein Blick auf die Zeilenprozentage zeigt sehr deutlich einen negativen Zusammenhang zwischen der Höhe des Katholikenanteils und den NSDAP-Stimmenanteilen: Der Prozentsatz der NSDAP-Stimmen bei der Reichstagswahl von 1933 fällt in den Gemeinden und Kreisen mit einer evangelischen Bevölkerungsmehrheit sehr viel größer aus als in Gemeinden und Kreisen mit einem höheren Katholikenanteil. Anders ausgedrückt: Es besteht ein negativer Zusammenhang zwischen dem Anteil katholischer Wähler und den Wahlerfolgen der Nationalsozialisten. Diese negative Beziehung kommt auch deutlich in den ordinalen Assoziationskoeffizienten zum Ausdruck:

Tab. 5.11: Katholikenanteil und NSDAP-Stimmenanteil bei der Reichstagswahl von 1933

Katholiken-Anteil	NSDAP-Anteil 11,4 %- 27,2 %	NSDAP-Anteil 27,3 %- 33,9 %	NSDAP-Anteil 34 %- 41,7 %	NSDAP-Anteil 41,8 %- 79,1 %	Zeilen-Summen
0 % - 33,33 %	5 4,1 %	18 14,8 %	34 27,9 %	65 53,3 %	122 100 %
33,34 % - 66,66 %	1 2,9 %	6 17,1 %	16 45,7 %	12 34,3 %	35 100 %
66,67 % - 100 %	99 37,4 %	82 30,9 %	56 21,1 %	28 10,6 %	172 100 %
Spalten-Summen	105 24,9 %	106 25,1 %	106 25,1 %	105 24,9 %	422 100 %

Tab. 5.12: Übersicht zu den Assoziationsmaßen bivariater Tabellenanalysen

Maßzahl	Skalenniveau der Merkmale	Tabellen- Größe	Assoziations- Modell	Werte- bereich
d %	nominal	2 x 2	Prozentsatz- differenz	0 %; +/- 100 %
- Phi - Kontingenz- koeffizient C - Cramer's V	nominal	2 x 2 k x m k x m	Abweichung von dem Modell der statistischen Unabhängigkeit	0; 1
- Kendall's Tau _a , Tau _b , Tau _c - Gamma - Somer's d	ordinal	k x m	Paarweiser Vergleich (ordinale Assoziation)	-1; +1
Lambda Eta ²	nominal nominal/ metrisch	k x m	Proportionale Fehlerreduk- tion	0; 1 (0, 100 %)

- Kendall's Tau-b: $-0,468$,
- Kendall's Tau-c: $-0,436$,
- Gamma : $-0,702$,
- Somer's d : $-0,565$ ($Y = \text{NSDAP-Stimmenanteil}$).

Gleich welchen Koeffizienten man zur Interpretation heranzieht: Man kommt in jedem Fall zu der Feststellung, daß eine starke negative Beziehung zwischen den kreuztabellierten Merkmalen »Katholikenanteil« und »NSDAP-Stimmenanteil« besteht. Wie man sieht, variieren die Maßzahlen in der Größenordnung etwas, und zwar zwischen gerundet $0,44$ und $0,56$. Nur Gamma weist einen deutlich höheren Wert auf, da diese Maßzahl keine Bindungen berücksichtigt. Gamma ist ein Assoziationsmaß, das im Sinne der proportionalen Fehlerreduktion interpretiert werden kann (ausführlich s. Benninghaus, S., a.a.O., S. 260ff). Die Vorhersageverbesserung beträgt hier 70% . Die Maßzahl Tau-c ist dem Tau-b-Koeffizienten vorzuziehen, da keine quadratische Tabelle vorliegt. Die asymmetrische Maßzahl Somer's d berücksichtigt die Bindungen in dem abhängigen Merkmal »NSDAP-Stimmenanteil« (T_y). Der numerische Wert ist geringfügig größer als Tau-c, da hier die Bindungen des unabhängigen Merkmals nicht berücksichtigt werden. Die Tabelle 5.12 stellt die behandelten Maßzahlen zur bivariaten Kontingenztabellenanalyse noch einmal zusammenfassend gegenüber.

5.6 Bivariate Modelle auf der Grundlage von Streudiagrammen

Allen bisher diskutierten Maßzahlen bivariater Assoziation zwischen nominal/ordinalen Merkmalen gemeinsam ist die zweidimensionale Häufigkeitsverteilung in Form einer Kontingenztafel. Die Berechnung der unterschiedlichen Maßzahlen erfolgt anhand der nach X und Y aggregierten Untersuchungseinheiten, d.h. anhand der Zellenhäufigkeiten. Weisen die Merkmale X und Y beide metrisches Skalenniveau auf, sind sie zunächst zu klassifizieren, um eine Kontingenztabellenanalyse durchführen zu können. Diese Vorgehensweise setzt die Wahl geeigneter Intervalle voraus. Klassifikation bedeutet allerdings einen Informationsverlust gegenüber den originären Werten. Im Extremfall kann für jede Untersuchungseinheit ein unterschiedlicher Wert bezüglich X und Y beobachtet werden. Es stellt sich somit die Frage nach einem statistischen Modell, das den Zusammenhang zweier metrischer Merkmale auf der Basis der originären Merkmalsausprägungen angibt.

Um einen ersten Eindruck über einen möglichen Zusammenhang zweier metrischer Merkmale zu gewinnen, fertigt man zunächst ein *Streudiagramm* (engl.: scatterplot) an, bestehend aus den Wertepaaren $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)$. Dazu wird das kartesische Koordinatensystem verwendet. Kann zwischen einem unabhängigen Merkmal und einem abhängigen Merkmal unterschieden werden, dann wird die horizontale Achse zur Repräsentation des unabhängigen Merkmals X verwendet, die vertikale Achse zur Repräsentation des

abhängigen Merkmals Y . Die Werte des einen Merkmals werden gegen die Werte des zweiten Merkmals in das Koordinatensystem eingezeichnet. Sämtliche Wertepaare ergeben dann Punkte in einer zweidimensionalen Ebene. Die Gesamtheit dieser Punkte wird als *Punktwolke* bezeichnet. Die Form der Punktwolke vermittelt einen visuellen Eindruck von der Beziehung zwischen den Merkmalen, bevor sie mit einer Maßzahl beschrieben wird. Aus dem Verlauf der Punktwolke lassen sich bereits Schlüsse über die Richtung der Beziehung ziehen, aus der Breite der Streuung der Punktwolke gewinnt man einen ersten Eindruck über die Stärke eines Zusammenhangs. Ist die Punktwolke sehr schmal, so liegt ein relativ starker Zusammenhang vor; ist die Punktwolke dagegen fast kreisförmig, dann liegt praktisch kein Zusammenhang vor. Zwischen zwei metrischen Merkmalen besteht ein positiver Zusammenhang, wenn (analog zum Fall ordinaler Merkmale) das Y -Merkmal tendenziell hohe Werte annimmt, wenn auch das Merkmal X hohe Werte annimmt und umgekehrt beide Merkmale gleichsinnig zu niedrigen Werten tendieren. Ein negativer Zusammenhang liegt vor, wenn hohe Y -Werte tendenziell mit niedrigen X -Werten auftreten. Die Abbildung 5.3 illustriert drei idealtypische Konstellationen von Streudiagrammen. In dem Streudiagramm (a) läßt sich kein Zusammenhang erkennen. Die Diagramme (b) und (c) zeigen dagegen jeweils eine gemeinsame Tendenz der X - und Y -Werte. In (b) gehen hohe Y -Werte mit hohen X -Werten einher und niedrige Y -Werte mit niedrigen X -Werten. Hier liegt ein positiver Zusammenhang vor. In (c) ist die Tendenz umgekehrt, d.h. dieses Diagramm illustriert einen negativen Zusammenhang.

Nehmen wir als überschaubares Beispiel die in der Tabelle 3.1b gegebenen Informationen über die Katholikenanteile in den 35 Wahlkreisen der Weimarer Republik und die NSDAP-Stimmenanteile von 1933. Die originären Merkmalsausprägungen sind in beiden Fällen metrisch. In der Abbildung 5.4 sind die dokumentierten Daten durch einzelne Wertepaare in einem Streudiagramm dargestellt. Im Rahmen der Tabellenanalyse mit klassifizierten Daten hatten wir auf der Ebene der Kreise und Gemeinden bereits einen negativen Zusammenhang zwischen diesen Merkmalen festgestellt. Eine Betrachtung der 35 Wahlkreise in dem Streudiagramm bestätigt den negativen Zusammenhang auch auf diesem hohen Aggregationsniveau. Mit steigendem Katholikenanteil beobachten wir deutlich niedrigere NSDAP-Stimmenanteile.

5.6.1 Korrelationsanalyse

Der auf die Statistiker August Bravais (1811–1863) und Karl Pearson (1857–1936) zurückgehende *Korrelationskoeffizient* r_{xy} beantwortet die Frage nach der Stärke eines linearen Zusammenhangs zwischen zwei metrischen Merkmalen X und Y . Um die Beziehung numerisch zu erfassen, wird die Lage eines jeden Wertepaares $(x_1, y_1), \dots, (x_n, y_n)$ mit Bezug zu dem Schwerpunkt (\bar{x}, \bar{y}) des Streudiagramms beschrieben. Legt man in das Streudiagramm ein Mit-

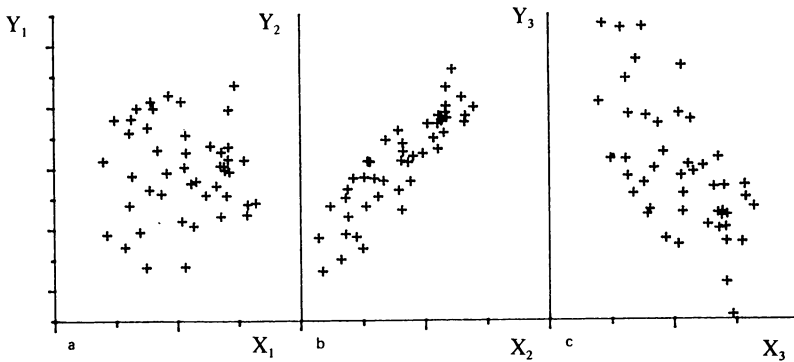


Abb. 5.3: Streudiagramme mit idealtypischen Punktwolken

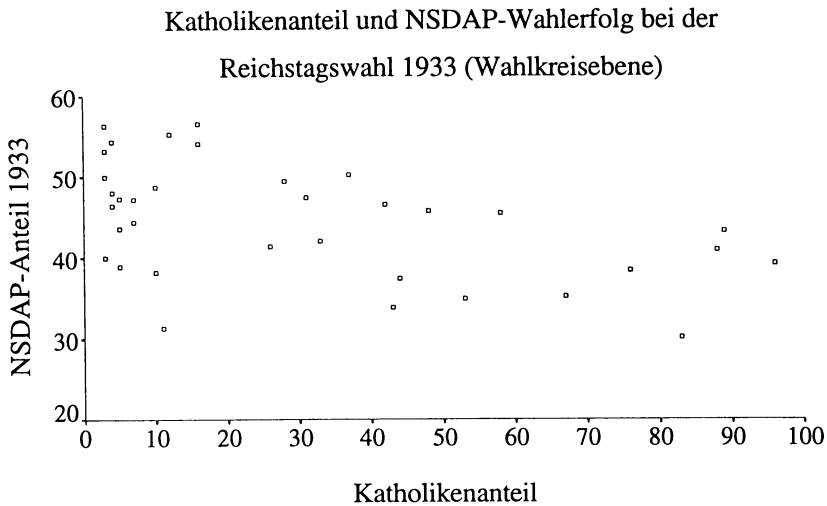


Abb. 5.4: Beispiel für ein Streudiagramm auf Wahlkreisebene

telwertkoordinatenkreuz, dann entstehen zwischen der X- und Y-Achse vier Quadranten (s. Abbildung 5.5). Nach der oben eingeführten Definition für die Richtung des Zusammenhangs deutet ein überwiegender Teil der Punkte im dritten und zweiten Quadranten auf einen positiven Zusammenhang hin; beobachtet man den überwiegenden Teil der Punkte im ersten und vierten Quadranten, so liegt ein negativer Zusammenhang vor. Formal wird dies für jedes Wertepaar im Streudiagramm durch das Produkt $(x_i - \bar{x})(y_i - \bar{y})$ erfasst. Dieses Produkt beschreibt, wie zwei metrische Merkmale miteinander *kovariieren*. Für jedes Wertepaar x_i, y_i ($i = 1, \dots, n$) gibt es vier mögliche Lagen im Streudiagramm:

Quadrant I : $x_i < \bar{x}$, $y_i > \bar{y}$; $(x_i - \bar{x})(y_i - \bar{y}) < 0$,
 Quadrant II : $x_i > \bar{x}$, $y_i > \bar{y}$; $(x_i - \bar{x})(y_i - \bar{y}) > 0$,
 Quadrant III: $x_i < \bar{x}$, $y_i < \bar{y}$; $(x_i - \bar{x})(y_i - \bar{y}) > 0$,
 Quadrant IV: $x_i > \bar{x}$, $y_i < \bar{y}$; $(x_i - \bar{x})(y_i - \bar{y}) < 0$.

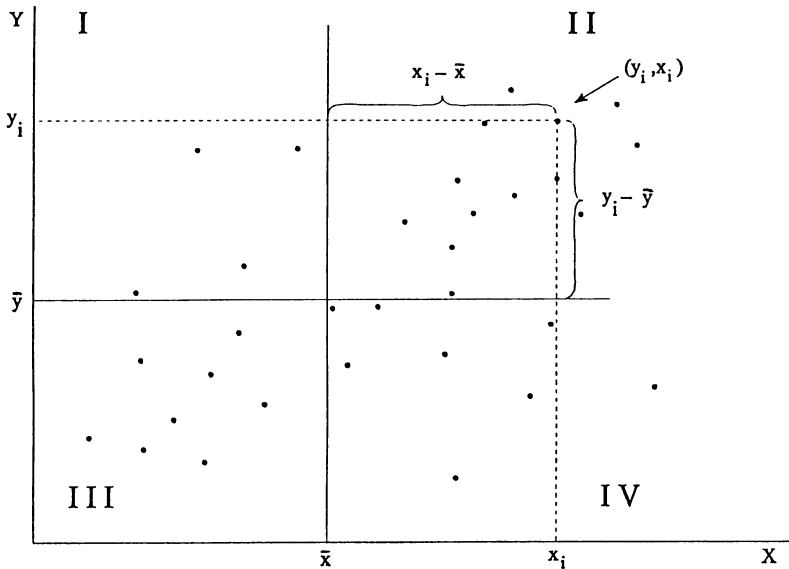


Abb. 5.5: Streudiagramm mit seinem Schwerpunkt

Liegen die Punkte hauptsächlich in den Quadranten II und III, so ist die Summe der Produkte stark positiv. Liegen die Punkte hauptsächlich in den Quadranten I und IV, so ist sie stark negativ. Sind die Punkte gleichmäßig über die vier Quadranten verteilt, so heben sich positive und negative Summanden weitgehend auf und die Summe der Produkte ist nahe Null. Um die Produktsumme von der Anzahl der n beobachteten Wertepaare unabhängig zu machen, wird diese durch n dividiert. Die Summe der Produkte $(x_i - \bar{x})(y_i - \bar{y})$, $i=1, \dots, n$, dividiert durch die Anzahl n wird als *Kovarianz* der Merkmale X und Y bezeichnet:

$$\text{cov}(x, y) = s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Als Maß für den Zusammenhang zwischen zwei metrischen Merkmalen X und Y ist die Kovarianz nicht ganz geeignet, da der numerische Wert der Kovarianz von den Maßeinheiten der Merkmale X und Y abhängig ist. Um die Skalenabhängigkeit zu beseitigen, wird die Kovarianz auf die Standardabweichungen von X (s_x) und Y (s_y) normiert. Das Ergebnis ist der Korrelationskoeffizient r nach Bravais – Pearson:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

$$= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Der Korrelationskoeffizient ist eine *standardisierte Kovarianz*, die als Verhältnis der Kovarianz zum Produkt der Standardabweichungen der beiden Variablen X und Y darstellbar ist. Nach Kürzung ergibt sich als vereinfachter Ausdruck:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Der Wert des Korrelationskoeffizienten liegt in dem Wertebereich -1 und $+1$. Je mehr sich der Wert der Maßzahl den Werten -1 bzw. $+1$ nähert, desto enger, intensiver ist der Zusammenhang. Ist der Korrelationskoeffizient gleich $+1$, so entspricht die Beziehung zwischen X und Y einer steigenden linearen Funktion, d.h. sämtliche Wertepaare liegen auf einer Geraden; ist die Korrelation gleich -1 , so entspricht sie einer fallenden linearen Funktion. Je mehr sich der Wert von r_{xy} dem Wert Null nähert, um so schwächer ist der untersuchte Zusammenhang. Sind die beiden untersuchten metrischen Merkmal überhaupt nicht linear verbunden, so ist $r_{xy} = 0$; es liegt dann keine lineare Korrelation vor. Die Abbildung 5.6 zeigt idealisierte X/Y-Streudiagramme zur Veranschaulichung unterschiedlich hoher Korrelationen (a: $r = +1$; b: $r = -1$; c: $r = +0.84$; d: $r = +0.33$).

Aus einem Korrelationskoeffizienten von Null kann nicht voreilig der Schluß gezogen werden, daß kein Zusammenhang zwischen den betrachteten Merkmalen besteht. Es kann lediglich der Schluß gezogen werden, daß keine lineare Korrelation feststellbar ist. Die Frage, ob eventuell eine nichtlineare Beziehung zu beobachten ist, kann mit Hilfe des Korrelationskoeffizienten nicht beantwortet werden. Auf diese Problematik werden wir in dem folgenden Abschnitt näher eingehen. Die Tabelle 5.13 enthält die Korrelationen zwischen dem Katholikenanteil und den NSDAP-Stimmenanteilen für die Reichstagswahlen 1930 bis 1933 auf Wahlkreisebene ($n = 35$).

Ab der Reichstagswahl 1930 gilt durchgängig, daß der Prozentsatz der NSDAP-Stimmen in Wahlkreisen mit einer katholischen Bevölkerungsmehrheit

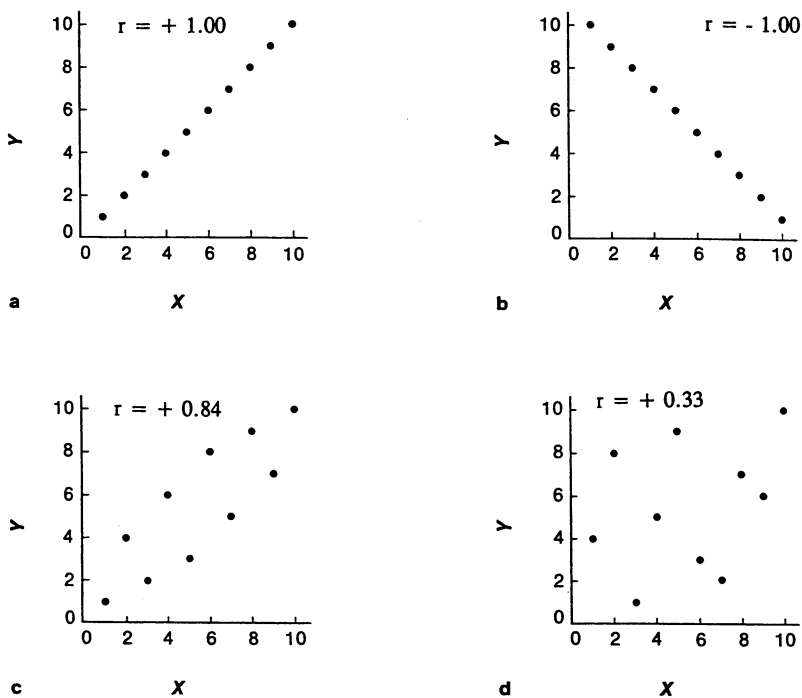


Abb. 5.6: Streudiagramme zur Veranschaulichung unterschiedlich hoher Korrelationskoeffizienten

Tab. 5.13: Korrelationen zwischen Katholikenanteil und NSDAP-Stimmenanteil von 1930 bis 1933

Konfessionsstruktur 1925	NSDAP-Stimmenanteile	Korrelationskoeffizient r_{xy}
Katholikenanteil	Reichstagswahl 14. September 1930	-0,54
Katholikenanteil	Reichstagswahl 31. Juli 1932	-0,71
Katholikenanteil	Reichstagswahl 6. November 1932	-0,64
Katholikenanteil	Reichstagswahl 5. März 1933	-0,49

sehr viel niedriger ausfiel als in Kreisen mit einem höheren Protestantenanteil. Es zeigt sich durchgehend ein starker negativer statistischer Zusammenhang zwischen dem Katholikenanteil und den Wahlerfolgen der NSDAP. Am stärk-

sten ausgeprägt ist dieses Ergebnis bei der Juliwahl 1932. Sicherlich fallen auf diesem hohen Aggregationsniveau auf Wahlkreisebene die Korrelationen numerisch besonders groß aus. Allerdings beobachten wir auf der Ebene der Kreise und Gemeinden ähnlich ausgeprägte Zusammenhänge zwischen diesen beiden Merkmalen (s. Falter, J., 1991: Hitlers Wähler, München, 175ff).

5.6.2 Lineare Einfachregression

Im Gegensatz zur Korrelationsanalyse, bei der die Merkmale symmetrisch betrachtet werden, beschäftigt sich die Regressionsanalyse¹ ausschließlich mit der Untersuchung von Abhängigkeiten zwischen metrischen Merkmalen. Soll ein Merkmal Y durch ein anderes Merkmal X erklärt werden, so ist die Einflußrichtung festgelegt. Die Aufgabe der Regressionsanalyse ist es, die Art der Abhängigkeit zu bestimmen, d.h. dasjenige mathematische Modell zu finden, durch das sich die zwischen den Merkmalen bestehende Abhängigkeit möglichst gut beschreiben läßt. Während bei der *Mehrfachregression* (der multiplen Regression) die Einflußbeziehungen zwischen zwei oder mehr Merkmalen bezüglich des abhängigen Merkmals Y untersucht werden, ist die *Einfachregression* nur auf die Untersuchung der Abhängigkeitsbeziehung zwischen zwei Merkmalen beschränkt. Die in dem Modell der Einfachregression zu erklärende Größe Y wird auch als Regressand oder Kriteriumsvariable bezeichnet und das unabhängige Merkmal X auch als Regressor- oder Prädiktorvariable. Die Modellbildung im Rahmen der Regressionsanalyse erfolgt in drei Schritten:

1. Zur modellgerechten Formulierung von asymmetrischen Merkmalsbeziehungen gehört die fachwissenschaftliche Festlegung des unabhängigen und abhängigen Merkmals (*Festlegung der kausalen Anordnung*). Ob ein Merkmal als abhängig oder unabhängig eingestuft wird, ergibt sich allein aus dem fachwissenschaftlichen oder sachlogischen Kontext.
2. Man spricht allgemein von einer Regression von Y auf X und nennt die mathematische Funktion, die die Abhängigkeit zwischen den beiden Merkmalen beschreibt, YX-Regressionsfunktion, oder formal ausgedrückt:

¹ Der Name Regression ist historisch bedingt durch das von Galton (1889) geprägte *Gesetz der universalen Regression*: »Each peculiarity in a man is shared by his kinsman but on the average in a less degree«, d.h. jede vom Normalen abweichende Eigenschaft eines Menschen wird von der nachfolgenden Generation zwar übernommen, aber durchschnittlich in einem geringeren Maße; bezüglich der Eigenschaft tritt also ein Rückschritt, eine Regression ein. Um dieses Gesetz empirisch zu prüfen, hat Galtons Freund Karl Pearson (1903) in 1078 Familien die Größe von Vater und Sohn untersucht. Dabei stellte er fest, daß – obwohl große Väter dazu neigen, große Söhne zu haben – die durchschnittliche Größe von Söhnen großer Väter geringer ist als die ihrer Väter: ist ein Vater um 1 cm größer als ein anderer, so ist sein Sohn um durchschnittlich 0,516 cm größer als der Sohn des anderen; umgekehrt läßt sich diese Aussage auf die Größe der Söhne kleiner Väter übertragen. Es ist also eine Regression in Bezug auf die Besonderheiten in der Größe bei den Söhnen zu erkennen.

$Y = f(X)$. Diese Gleichung zeigt die allgemeinste Form einer Einflußbeziehung zwischen den Merkmalen X und Y . Fachwissenschaftliche Hypothesen oder Vermutungen liefern in der Regel nur Aussagen zur Qualifizierung des abhängigen und unabhängigen Merkmals, in den seltensten Fällen dagegen konkrete Hinweise über die mathematische Form der Beziehung. Im allgemeinen kann man Hinweise über die mathematische Form der Regressionsfunktion nur aus empirischen Beobachtungswerten der Merkmale gewinnen, z.B. durch die Inspektion der Punktwolke in einem zweidimensionalen Streudiagramm. Aus der Punkteverteilung lassen sich erste Hinweise auf einen zur Beschreibung der Beziehung zwischen X und Y geeigneten Funktionstyp ableiten. Einführend soll in diesem Abschnitt nur auf die einfache *lineare Regression* eingegangen werden. Zum einen hat sich gezeigt, daß viele Abhängigkeiten zwischen Merkmalen zumindest näherungsweise recht gut durch lineare Funktionen dargestellt werden können. Zum anderen ist es in vielen Fällen möglich, nichtlineare Funktionen zu linearisieren, d.h. durch geeignete Merkmalstransformationen in lineare Funktionen zu überführen.

3. Jede Modellbildung ist schließlich noch auf die Beurteilung der Güte der Modellanpassung an den empirischen Befund angewiesen. Kriterien für die Güte der Anpassung beruhen auf Maßzahlen, die sich aus dem bereits bekannten PRE-Modell ableiten lassen.

Grundsätzlich gibt es zwei Arten von Abhängigkeitsbeziehungen der allgemeinen Form $Y = f(X)$:

- Eine mathematische lineare Funktion, die den Zusammenhang zwischen zwei Größen exakt, d.h. *deterministisch* spezifiziert und
- eine *stochastisch* spezifizierte lineare Funktion.

Die denkbar einfachste Beziehung zwischen zwei metrischen Größen ist die perfekt lineare mathematische Funktion wie z.B. $Y = 1x$ oder $Y = 5 - 2x$. Wie die Beispiele in der Abbildung 5.7 zeigen, kann eine perfekt lineare Beziehung geometrisch als Gerade und algebraisch als lineare Gleichung dargestellt werden. Diese Beispiele sind spezielle Fälle der generellen Geradengleichung

$$y = a + bx .$$

Der *Steigungskoeffizient* b gibt an, um wieviel Einheiten sich Y verändert, wenn X um eine Einheit zunimmt. Der *Achsenabschnitt* a zeigt an, welchen Wert Y an der Stelle $X = 0$ annimmt (soweit eine solche Interpretation überhaupt sinnvoll ist). Das Vorzeichen des Steigungskoeffizienten drückt die Richtung der Abhängigkeit aus, d.h. ob eine Beziehung positiv oder negativ ist.

Eine stochastische Abhängigkeit liegt vor, wenn bei gegebenen Werten des unabhängigen Merkmals X die Werte des abhängigen Merkmals in einem mehr oder weniger großen Intervall streuen. Auch wenn zwischen den beiden Merkmalen eine theoretisch gültige Einflußbeziehung besteht, werden die Beobach-

tungswerte des Merkmals Y nie direkt auf der Geraden liegen, sondern bei gegebenen X-Werten mehr oder weniger um diese herum streuen. Diese Streuungen können u.a. dadurch bedingt sein, daß die abhängige Größe Y noch von weiteren Einflußfaktoren abhängt, die aber unberücksichtigt bleiben. Ferner kann davon ausgegangen werden, daß die Messungen der Beobachtungswerte nicht exakt vorgenommen wurden. Mit Streuungserscheinungen (gleich welcher Ursache) entsteht die Notwendigkeit, die mathematische Funktion stochastisch zu formulieren. Da bei gleichem X-Wert verschiedene Y-Werte auftreten können, geht man von einem nichtdeterministischen Modellansatz aus, der das Prinzip der stochastischen Abhängigkeit zu erkennen gibt:

$$Y = f(X, E) .$$

Mit dem Symbol »E« (engl.: Error) werden im folgenden sämtliche Einflüsse auf Y belegt, die nicht aus dem explizit berücksichtigten Merkmal X herrühren, aber dennoch auf das Merkmal Y einwirken. Die Fehlervariable E (auch als *Störgröße* bezeichnet) repräsentiert auch Abweichungen vom unterstellten linearen Zusammenhang. Während für das unabhängige und abhängige Merkmal gemeinsame Beobachtungswerte vorliegen (Wertepaare y_i, x_i), ist die Störgröße nicht direkt beobachtbar, da sie ein Konglomerat vieler verschiedener Einflußfaktoren darstellt.

Ausgangspunkt für ein bivariates lineares Regressionsmodell ist wiederum ein Streudiagramm, bestehend aus den Wertepaaren $(x_i, y_i; i = 1, \dots, n)$, die die Punktwolke bilden. Das Ziel besteht darin, eine geeignete stochastische Funktion $f(X, E)$ anzugeben, die die Abhängigkeit der Variablen Y von X beschreibt. Graphisch gesehen lautet die Aufgabe, eine geeignete Gerade durch die Punkte $(x_i, y_i), i = 1, \dots, n$ des Streudiagramms zu legen. Die Ausgangsgleichung erhält nun für jedes Wertepaar die Form:

$$y_i = a + bx_i + e_i, \quad i = 1, \dots, n .$$

Die Regressionsgleichung enthält auf der rechten Seite neben der Konstanten a die *systematische Komponente* bx_i und den Term e_i , der als *Residuum* bezeichnet wird. In dem Residuum repräsentiert die Störgröße E; sie umfaßt sämtliche »restlichen« Einflüsse auf y_i (fehlende erklärende Merkmale, Meßfehler, Abweichungen von der Linearitätsannahme) erfaßt. Die Abbildung 5.8 stellt vergleichend (a) eine deterministische lineare Beziehung und (b) eine stochastische Beziehung gegenüber.

Gesucht wird nun die Regressionsgerade, die den Verlauf der Y/X-Punktwolke bestmöglich beschreibt. Die Funktions- oder Schätzwerte auf der zu bestimmenden Geraden werden mit einem »Dach« über dem y_i -Wert symbolisiert. Zu jedem Beobachtungswert x_i ($i = 1, \dots, n$) läßt sich der zugehörige Schätzwert auf der Geraden angeben:

$$\hat{y}_i = f(x_i), \quad \hat{y}_i = a + bx_i, \quad i = 1, \dots, n .$$

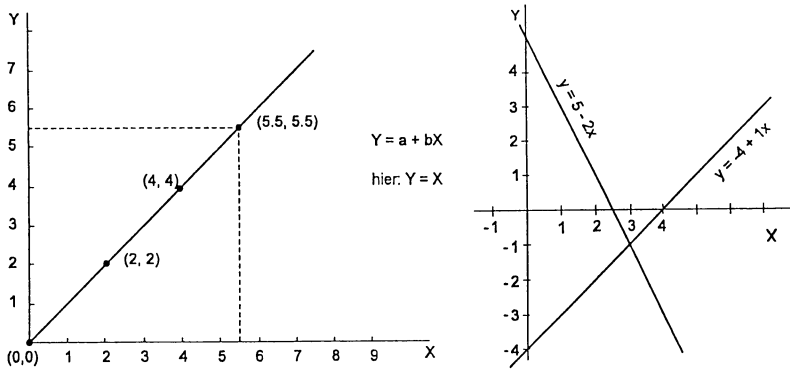
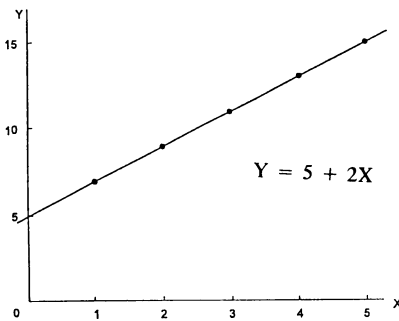


Abb. 5.7: Beispiele für lineare Funktionen

(a) Deterministisches Modell: Perfekte Geradengleichung



(b) Statistisches Modell: Streuung oberhalb und unterhalb der Geraden

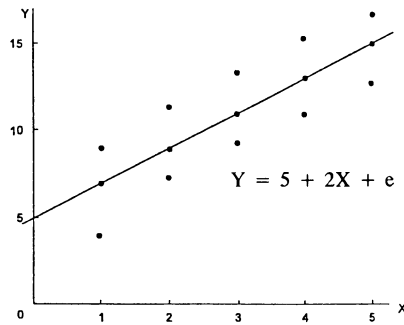


Abb. 5.8: Deterministisches und stochastisches Regressionsmodell

Die Grundidee zur Bestimmung der optimalen Regressionsgeraden besteht darin, die Abweichungen zwischen den beobachteten y_i -Werten und den Schätzwerten \hat{y}_i auf der Geraden zu minimieren. Formal lassen sich diese Abweichungen e_i beschreiben durch die Gleichung

$$e_i = y_i - \hat{y}_i .$$

Damit ist die Stör- oder Residualgröße eindeutig definiert. Die Abbildung 5.9 illustriert das Verhältnis von Beobachtungswert, Schätzwert auf der Regres-

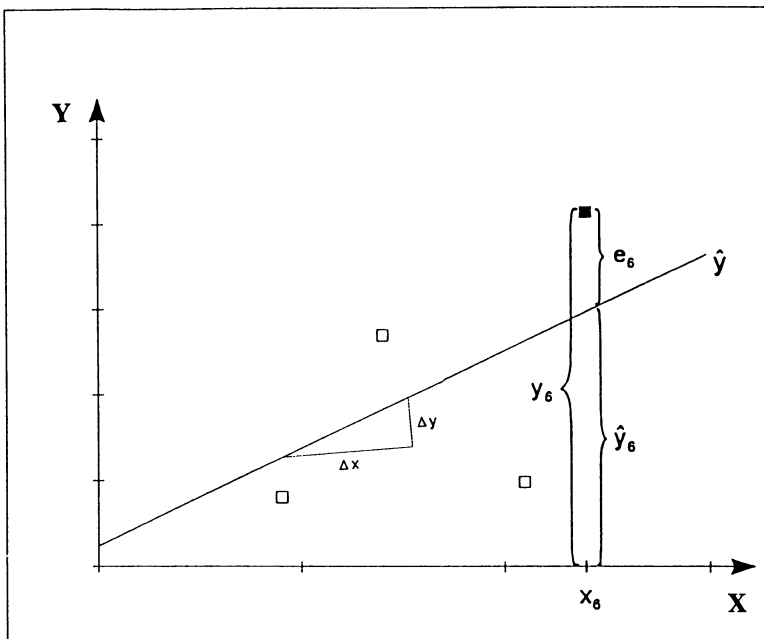


Abb. 5.9: Hypothetische Regressionsgerade und Residualgröße

sionsgeraden und Residualgröße als vertikale Streckenzerlegung im Streudiagramm für einen X-Wert der hypothetischen Regressionsgeraden.

Das Regressionsmodell läßt sich somit beschreiben als eine additive Zerlegung jeder Beobachtung y_i ($i = 1, \dots, n$) in eine systematische Komponente, die sich linear mit den X-Werten verändert, und der Residualgröße e_i (s. Tabelle 5.14).

Tab. 5.14: Additive Zerlegung der Beobachtungswerte des abhängigen Merkmals

Werte des abhängigen Merkmals	Systematische Komponente	Störgröße oder Residuum
$y_i, i = 1, \dots, n$	$\hat{y}_i, i = 1, \dots, n$	$e_i, i = 1, \dots, n$
$y_i =$	$a + bx_i$	$+ e_i$

Die Größenveränderung eines abhängigen Merkmals wird bei der linearen Einfachregression auf den Einfluß eines theoretisch spezifizierten unabhängigen Merkmals zurückgeführt. Der Funktionsverlauf wird durch das Vorzeichen des Regressionskoeffizienten bestimmt, die Regressionskonstante (a) drückt als Ordinatenschnittpunkt das »Niveau« der Geraden aus und der Steigungskoeff-

fizient (b) beschreibt den Wertezuwachs in dem abhängigen Merkmal, wenn das unabhängige Merkmal um eine Maßeinheit vergrößert wird.

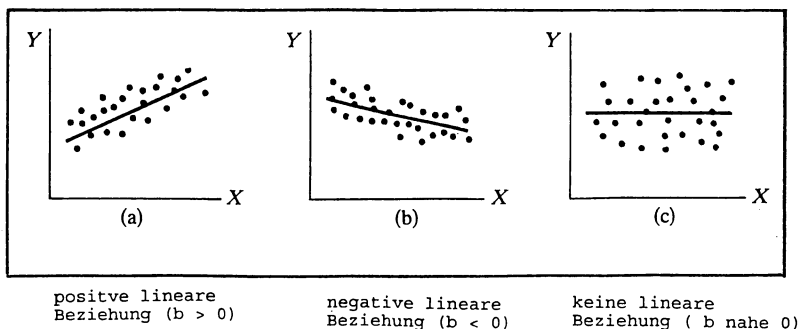


Abb. 5.10: Idealtypische Geradenverläufe

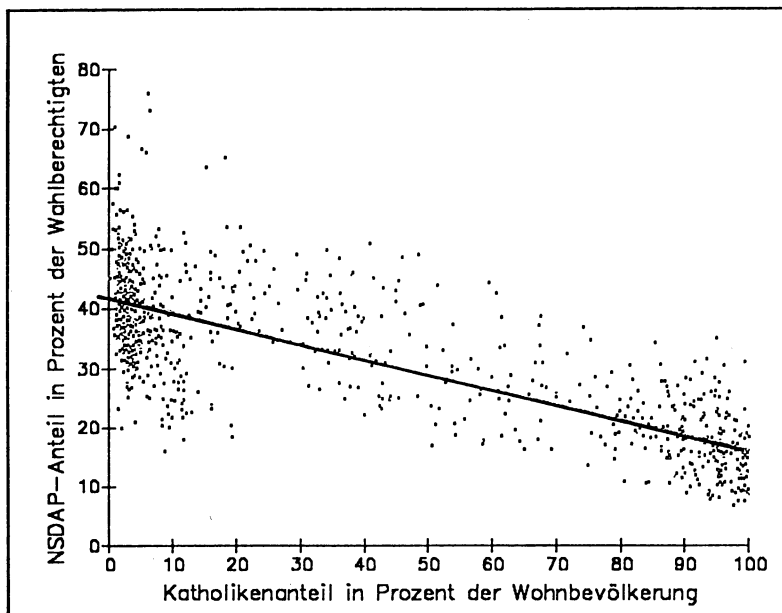


Abb. 5.11: Katholikenanteil und NSDAP-Wahlerfolg bei der Reichstagswahl vom Juli 1932 (Freihand-Regressionsgerade)

Die Abbildung 5.10 verdeutlicht unterschiedliche Geradenverläufe in exemplarischen Streudiagrammen. Die allgemeine Tendenz der Punktwolke weist im Diagramm (a) nach rechts oben und lässt sich demnach durch eine Gerade mit positivem Anstieg beschreiben. In dem Diagramm (b) weist die allgemeine Tendenz der Punktwolke nach rechts unten. In diesem Fall liegt eine negative

Beziehung vor. Das Diagramm (c) illustriert den Extremfall völliger Zusammenhangslosigkeit. In diesem Fall verläuft die Gerade parallel zur X-Achse. Die Y-Werte streuen »zufällig« um die eingezeichnete Horizontale.

Das folgende Beispiel zeigt die bivariate Abhängigkeit zwischen den Stimmenanteilen der NSDAP bei der Reichstagswahl im Juli 1932 und dem Prozentanteil der Katholiken auf der Ebene der Kreise des deutschen Reiches. Das Streudiagramm in der Abbildung 5.11 verdeutlicht, daß der NSDAP-Anteil umso höher ist, je niedriger der Prozentsatz der Katholiken ist. Durch diese Punktwolke läßt sich eine Gerade mit einem negativen Steigungswinkel legen. Zeichnet man in diese Punktwolke »freihändig« mit einem Lineal nach bestem subjektiven Augenmaß eine »optimale« Gerade ein, dann kann der Wert für a (der Ordinatenabschnitt) und b (in diesem Beispiel die Verminderung des NSDAP-Anteils bei einer Veränderung des Katholikenanteils um eine Prozenteinheit) dem Diagramm direkt entnommen werden. Die Freihand-Regressionsgerade schneidet die Y-Achse bei einem NSDAP-Anteil von 42% (Regressionskonstante a), der Steigungskoeffizient b beträgt 0,25. Die Ergebnisse der Freihand-Regression können nun leicht in »Wenn-Dann-Sätze« transformiert werden. Wir können mit Hilfe des Steigungskoeffizienten folgende Aussage treffen: Wenn sich der Katholikenanteil um 1% erhöht, dann vermindert sich der geschätzte Stimmenanteil der NSDAP bei der Reichstagswahl im Juli 1932 um 0,25%. Die geschätzten NSDAP-Anteile liegen auf der Regressionsgeraden $NSDAP(1932)\% = 42 - 0,25 \cdot KATH$. Man kann die durch eine Regressionsgerade beschriebene Abhängigkeit auch zu Vorhersagen verwenden. Einem Katholikenanteil von z.B. 50% entspricht ein geschätzter NSDAP-Stimmenanteil von 29,5% ($= 42\% - 0,25 \cdot 50\%$). Die tatsächlichen NSDAP-Stimmenanteile streuen für diesen Katholikenanteil oberhalb und unterhalb der Geraden, d.h. die empirischen Werte für das abhängige Merkmal sind bei $x = 50\%$ sowohl größer als auch kleiner als der geschätzte Wert von 29,5%. Wir wollen dieses Beispiel nicht weiter vertiefen und uns im folgenden den rechnerischen Problemen der Bestimmung von a und b zuwenden, denn mit einer subjektiv gefundenen Freihandgeraden können wir in der Regel nicht auskommen. Die »optimale« Lage der Geraden in der Punktwolke des Streudiagramms folgt aus objektiven, d.h. mathematischen Kriterien.

Durch die Wertepaare $(y_i, x_i; i = 1, \dots, n)$ soll eine Gerade gelegt werden, die sich möglichst gut an diese Punkte anpaßt. Damit ist gemeint, daß sämtliche Wertepaare einer Punktwolke möglichst dicht an der Geraden liegen sollten. Der Terminus »möglichst dicht« ist im folgenden mathematisch zu präzisieren. Das Problem konzentriert sich darauf, die unbekannten Koeffizienten a und b auf der Grundlage eines formalen Optimalitätskriteriums unter Verwendung der Beobachtungswerte des unabhängigen und abhängigen Merkmals numerisch zu bestimmen. Als Ausgangspunkt können die Residuen $(e_i = y_i - \hat{y}_i; i = 1, \dots, n)$ dienen. Möglichst gut beschreibt eine Gerade die Punktwolke, wenn diese Abweichungen insgesamt möglichst klein sind. Es gibt verschiedene formale An-

sätze, die Abweichungen zusammenzufassen. Die Summe der Abweichungen ist gleich Null, da die Differenzen unterschiedliche Vorzeichen aufweisen und sich so gegenseitig aufheben. Als Kriterium ist die Summe der Abweichungen nicht ausreichend, da es praktisch unendlich viele Geraden gibt, die diese Bedingung erfüllen, und zwar all jene Geraden, die durch den Schwerpunkt des Koordinatensystems (\bar{x}, \bar{y}) gehen. Daher ist eine zweite Bedingung notwendig, die sicherstellt, daß sich positive und negative Abweichungen nicht gegeneinander aufheben. Das Kriterium der *kleinsten absoluten Abweichungen* fordert z.B., daß a und b so bestimmt werden, daß die Summe der absoluten Abweichungen gleich ein Minimum wird:

$$\sum_{i=1}^n |(y_i - (\hat{a} + \hat{b} x_i))| \Rightarrow \text{Min!}$$

Dieses Prinzip ist – historisch gesehen – das wohl älteste objektive Anpassungskriterium für eine Regressionsgerade. Wegen der beträchtlichen formalen Schwierigkeiten, die Koeffizienten a und b nach diesem Ansatz zu bestimmen, trat dieses Kriterium praktisch völlig in den Hintergrund. Eine weitere Möglichkeit besteht darin, die zu minimierende Summe aus den quadrierten Differenzen zu bilden. Dieses Kriterium wird *Methode der kleinsten Quadrate* genannt oder *OLS-Schätzung* (engl.: ordinary least squares) und lautet:

$$\sum_{i=1}^n (y_i - (\hat{a} + \hat{b} x_i))^2 \Rightarrow \text{Min!}$$

Die y_i - und x_i -Werte in der zu minimierenden Funktion sind als beobachtete Werte konstante Größen, die Regressionskoeffizienten a und b sind die Veränderlichen. Es handelt sich hier somit um das Problem der Extremwertbestimmung einer Funktion mit zwei Veränderlichen. Eine notwendige Bedingung für das Vorliegen eines Extremwertes besteht darin, daß die beiden ersten partiellen Ableitungen Null werden. Zur Bestimmung des Minimums von $\Sigma(y_i - (a + bx_i))^2$ sind somit die partiellen Ableitungen für die beiden Unbekannten a und b zu bilden und gleich Null zu setzen. Hierdurch erhält man folgendes Gleichungssystem:

$$\begin{aligned} \frac{\delta}{\delta a} \sum_{i=1}^n (y_i - a - bx_i)^2 &= -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\delta}{\delta b} \sum_{i=1}^n (y_i - a - bx_i)^2 &= -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \end{aligned}$$

Nullsetzen der beiden partiellen Ableitungen und Auflösung der Summen führt zu den sogenannten *Normalgleichungen* der Regressionsanalyse:

1. Normalgleichung:

$$\sum_{i=1}^n y_i = n a + b \sum_{i=1}^n x_i$$

2. Normalgleichung:

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2$$

Die Normalgleichungen bilden gewissermaßen den Angelpunkt der Kleinst-Quadrate-Methode zur Bestimmung der Regressionsgeraden, denn die Koeffizienten a und b minimieren die Funktion dann und nur dann, wenn sie die Normalgleichungen erfüllen. Die zwei Unbekannten lassen sich nun einfach als Funktion von n , x_i und y_i angeben. Aus der ersten Normalgleichung folgt durch Division mit n :

$$\text{Wegen } \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}: \quad \bar{y} = a + b\bar{x}$$

$$a = \bar{y} - b\bar{x}$$

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$$

Aus der zweiten Normalgleichung erhält man die Lösung für den Steigungskoeffizienten:

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

Zum Schluß sei noch eine wichtige Abwandlung dieses Ausdrucks angegeben. Werden anstelle der Beobachtungswerte (x_i, y_i) die Abweichungen vom Mittelwert verwendet, vereinfacht sich die Berechnung. Für den Steigungskoeffizienten b ergibt sich dann:

Alternativformel:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Der Regressionskoeffizient b läßt sich somit darstellen als Verhältnis aus der Kovarianz von X und Y und der Variation von X ($b = s_{xy}/s_x^2$).

In der einfachen linearen Regressionsanalyse spielen vier Variablen eine zentrale Rolle:

1. die abhängige Variable y_i (das Kriterium),
2. die unabhängige Variable x_i (der Prädiktor),
3. die Schätzvariable \bar{y}_i (Modellvorhersage) und
4. die Störgröße $e_i = y_i - \bar{y}_i$ (Restvariable oder Residuum).

Wir wollen im folgenden ausgewählte Eigenschaften dieser vier regressionsanalytischen Variablen skizzieren und bestimmte wechselseitige Beziehungen hervorheben. Damit werden wesentliche Eigenschaften des regressionsanalytischen Modells deutlicher. Auf eine mathematische Ableitung wollen wir hier verzichten (vgl. z.B. Gaensslen, H./Schubö W., 1976: Einfache und komplexe statistische Analyse, 2. verb. A., München/Basel, S. 23ff).

- 1) Die Summe der Residuen e_i ($i = 1, \dots, n$) ist gleich Null.
- 2) Die Summe der Produkte $x_i \cdot e_i$ ($i = 1, \dots, n$) ist gleich Null.
- 3) Die Regressionsgerade geht durch den Schwerpunkt (\bar{x}, \bar{y}) der Punktwolke und damit gilt $\bar{y} = a + b\bar{x}$.
- 4) Das arithmetische Mittel der beobachteten Y -Werte ist gleich dem arithmetischen Mittel der geschätzten \bar{Y} -Werte.
- 5) Kovarianz und Korrelation des unabhängigen Merkmals X mit der Störgröße E sind gleich Null.
- 6) Kovarianz und Korrelation der Schätzvariable mit der Störgröße sind gleich Null.

Die einfache lineare Regressionsanalyse ist somit ein Modell, das die vorherzusagende abhängige Variable so in die Schätz- und Restvariable aufspaltet, daß die Schätzung die Unterschiedlichkeit der Untersuchungseinheiten in Y ausdrückt, die auf die Unterschiedlichkeit in der unabhängigen Variablen X zurückgeht.

Unter der Voraussetzung, daß die Beschreibung der Punkte (x_i, y_i) durch eine Gerade überhaupt sinnvoll ist, stellt sich die Frage, wie gut die Gerade den empirischen Befund beschreibt. Es ist daher noch eine Maßzahl für den Grad

der Modellanpassung einer Regressionsfunktion an die beobachteten Y-Werte anzugeben. Ein naheliegendes Kriterium für die Güte der Modellanpassung wären z.B. die Residuen. Analog zur Varianz des Merkmals Y

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

kann auch die Varianz der Residuen bestimmt werden:

$$s_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Je nachdem, wie weit die beobachteten Y-Werte um die geschätzte Regressionsgerade streuen (bzw. variieren), ergeben sich unterschiedliche gute Modellanpassungen an den beobachteten empirischen Befund. Die Streudiagramme in Abbildung 5.12 verdeutlichen unterschiedlich gute Modellanpassungen. Diagramm (a) zeigt eine perfekte Modellanpassung, Diagramm (b) eine mittlere und Diagramm (c) eine geringe Güte der Modellanpassung.

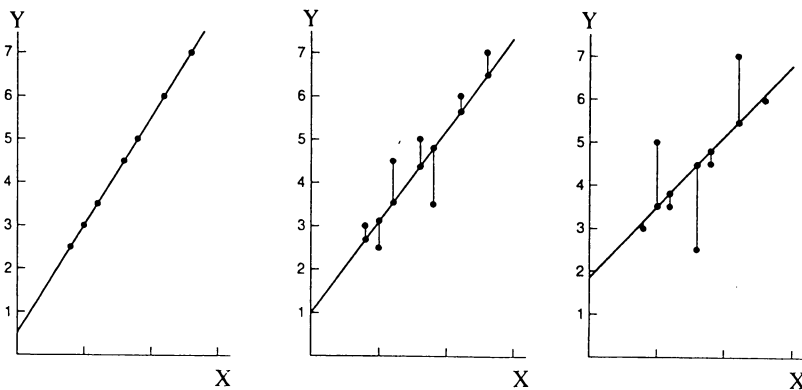


Abb. 5.12: Illustration unterschiedlicher Modellanpassungsgüten

Während die Varianz des abhängigen Merkmals s_y^2 die Differenz zwischen den beobachteten Werten y_i und dem arithmetischen Mittel von Y quantifiziert, bestimmt die Varianz der Residuen s_e^2 , in welchem Umfang die beobachteten Y-Werte um die Gerade variieren. Mit anderen Worten: Die Varianz s_e^2 ist ein Ausdruck für die verbleibende Varianz in Y, nachdem ein Teil der Ursprungsvarianz s_y^2 durch ein Regressionsmodell verringert, d.h. »erklärt« ist. Wir erwarten somit eine deutliche Wertedifferenz zwischen s_y^2 und s_e^2 . Diese Größen haben zwar eine untere Grenze von Null, eine obere Grenze kann allerdings nicht angegeben werden. Dennoch liefert ein Vergleich der beiden Varianzen

einen ersten Hinweis für die Güte der Modellanpassung. Als Maßzahl läßt sich dieser Varianzvergleich noch nicht heranziehen, da der absolute Wert der Differenz von der Größe der Maßeinheit des abhängigen Merkmals Y abhängt. Um zu einem solchen Maß zu gelangen, sind folgende –dem PRE-Konzept der Varianzanalyse analoge – Betrachtungen erforderlich.

- 1) Wenn lediglich die univariate Verteilung von Y betrachtet wird, dann ist die Antwort auf die Frage nach der besten Vorhersage der Y-Werte (ohne X-Variable) das arithmetische Mittel von Y. Die Varianz des Vorhersagefehlers ist in diesem Fall gleich der Varianz von Y (s_y^2), die zur Unterscheidung jetzt als *Gesamtstreuung* bezeichnet wird.

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Das arithmetische Mittel kann auch als Konstante in das Streudiagramm gelegt werden. Als Regressionsgerade verläuft sie dann horizontal zur X-Achse im Abstand von \bar{y} . Dieses Ausgangsmodell bezeichnet man auch als *Basismodell*.

- 2) Das zweite Modell (das sogenannte *aktuelle Modell*) unter Einbeziehung des unabhängigen Merkmals X wird eine mehr oder weniger große Vorhersage-Verbesserung gegenüber dem Basismodell bedeuten. Um beurteilen zu können, wie gut die Verbesserung der Vorhersage aufgrund von X gegenüber dem Basismodell ausfällt, kann man die ursprüngliche Gesamtstreuung in zwei Anteile zerlegen. Einen Anteil, der auf den Einfluß von X zurückzuführen ist, berechnet als Varianz der Abweichungen der geschätzten Y-Werte vom arithmetischen Mittel von Y (»erklärte Streuung« genannt):

$$s_{\hat{y}}^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

und einen Anteil, der auf die Störgröße zurückzuführen ist, berechnet als Varianz der Residuen (»nicht erklärte Streuung« genannt). Sie vertritt in dem Modellansatz sämtliche nicht berücksichtigten unabhängigen Merkmale (aber auch die Meßfehler und die Abweichungen von der Linearitätsannahme!):

$$s_e^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2 = \frac{1}{n} \sum_{i=1}^n (e_i - 0)^2 = \frac{1}{n} \sum_{i=1}^n e_i^2$$

$$s_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Diese drei Komponenten ergeben die *Varianzzerlegungsformel* des Regressionsmodells. Sie besagt, daß sich die Gesamtvarianz der Y-Werte aus zwei Komponenten zusammensetzt: Aus der Streuung der Werte um die Regressionsgerade und aus der Streuung, die durch die Schätzwerte auf der Regressionsgeraden bewirkt wird, d.h. einer Streuung, die allein aus der Steigung der Regressionsgeraden hervorgeht:

$$s_y^2 = s_y^2 + s_e^2$$

Die Abbildung 5.13 illustriert die grundlegende Logik dieser Zerlegung, die aus dem Übergang vom Basismodell ohne X zu einem aktuellen Modell mit X folgt, anhand von drei Wertepaaren in einem hypothetischen Streudiagramm. Die Abbildung (a) verdeutlicht die Abweichungen der Y-Werte vom arithmetischen Mittel des Merkmals Y (Gesamtstreuung); Die Abbildung (b) zeigt die Abweichungen der beobachteten Y-Werte von den Schätzwerten auf der Geraden; die Abbildung (c) zeigt die Abweichungen der Schätzwerte auf der Geraden von dem arithmetischen Mittel des Merkmals Y (als Abstand des aktuellen Modells zum Basismodell). Die Abbildung (c) zeigt die additive Verknüpfung der erklärten und nicht erklärten Anteile zur Gesamtabweichung als Zusammenfassung der drei vertikalen Streckenzerlegungen.

Aus der Streuungszerlegung läßt sich eine PRE-Maßzahl ableiten, die analog zum ETA^2 der Varianzanalyse als Anteil erklärter Varianz interpretiert werden kann. Das *Bestimmtheitsmaß* (symbolisiert mit $R^2_{y,x}$), das auch als *Determinationskoeffizient* bezeichnet wird, folgt aus dem Verhältnis der erklärten Varianz zur Gesamtvarianz. Diese Maßzahl drückt die Güte der Beschreibung der Punkte im Streudiagramm durch die Regressionsgerade aus. Die generelle Formel zur Berechnung der proportionalen Fehlerreduktion in Form des Bestimmtheitsmaßes $R^2_{y,x}$ lautet dann:

$$R^2_{y,x} = \frac{E_1 - E_2}{E_1} = \frac{\text{Gesamtvarianz} - \text{Nicht erklärte Varianz}}{\text{Gesamtvarianz}} = \frac{\text{Erklärte Varianz}}{\text{Gesamtvarianz}}$$

Unter Verwendung der oben verwendeten Varianznotation erhalten wir:

$$R^2_{y,x} = (s_y^2 - s_e^2) / s_y^2 = s_y^2 / s_y^2$$

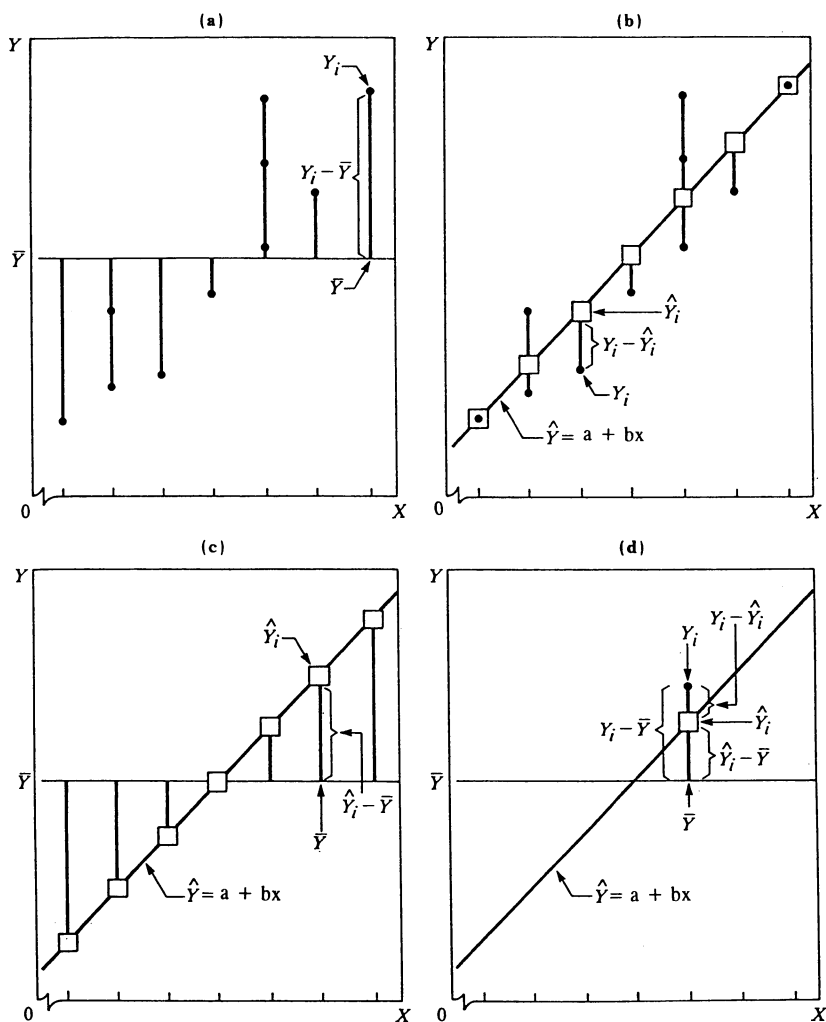


Abb. 5.13: Streuungszerlegung im Regressionsmodell

Quelle: Neter, J./Wasserman, W./Kutner, M.H., 1989: Applied Linear Regression Models, 2nd. edition, Boston, S. 88.

Die Division durch n lässt sich in den Varianzausdrücken herauskürzen. Das Bestimmtheitsmaß folgt dann allein aus den einzelnen Abweichungsquadratsummen der Variationsquellen. Damit erhält man eine Zerlegung der zu

erklärenden Gesamtabweichungsquadratsumme in die nicht erklärte Quadratsumme und die durch die Regressionsgerade erklärte Abweichungsquadratsumme. Das Bestimmtheitsmaß ist dann definiert als der Anteil der erklärten Variation (= Gesamtvariation – Nicht erklärte Variation) dividiert durch die Gesamtvariation des abhängigen Merkmals:

$$R^2 = \frac{E_1 - E_2}{E_1} = \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$= \frac{\text{Gesamtvariation} - \text{Nicht erklärte Variation}}{\text{Gesamtvariation}}$$

Das Bestimmtheitsmaß kann Werte zwischen 0 und 1 annehmen und gibt – mit 100 multipliziert – an, wieviel Prozent der Streuung der beobachteten Y-Werte durch die Regressionsgerade erklärt werden. Wenn sämtliche empirischen Werte y_i auf der Regressionsgeraden liegen, nimmt das Bestimmtheitsmaß den Wert Eins an. Wenn das Bestimmtheitsmaß gleich Null ist, sind sämtliche Schätzwerte \hat{y}_i gleich dem arithmetischen Mittel \bar{y} . Die Regressionsgerade ist dann eine Parallele zur X-Achse in der Höhe von \bar{y} . Die Regressionsfunktion hat in diesem Fall einen Regressionskoeffizienten von b gleich Null und keinerlei Erklärungswert. Das heißt nicht zwingend, daß zwischen X und Y keine Abhängigkeit besteht, denn es kann durchaus eine nichtlineare Beziehung vorliegen. In der Abbildung 5.14 sind idealtypische Punktwolken mit unterschiedlichen R^2 -Werten gegenübergestellt.

Der Determinationskoeffizient gibt Auskunft darüber, wieviel Varianz von Y mit Hilfe eines linearen Regressionsmodells erklärt werden kann, wobei »Erklärung« allein im statistischen Sinne zu verstehen ist. Der zugrunde liegende Modellgedanke meint die Verbesserung der Modellanpassungsgüte mit Hilfe der linearen Regressionsfunktion (gleich dem aktuellen Modell mit der Prädiktorvariablen X) gegenüber einer Schätzung mit Hilfe des Mittelwertmodells für die Prognose des abhängigen Merkmals Y (gleich dem Basismodell ohne die Prädiktorvariable X). Je größer der Teil der Gesamtvarianz ist, der durch die berechnete Regressionsfunktion erklärt werden kann, desto besser ist die Regressionsgerade den empirischen Wertepaaren in der Punktwolke angepaßt.

Die Abbildung 5.15 zeigt beispielhaft zwei Streudiagramme mit den unabhängigen Merkmalen X_1 und X_2 . Beide Punktwolken lassen sich durch lineare Funktionen mit numerisch gleichen Regressionskoeffizienten beschreiben (d.h. $a_1 = a_2$ und $b_1 = b_2$). Die erste Regressionsgerade »schmiegt« sich sehr gut an die Wertepaare der Punktwolke an, während die zweite Regressionsgerade eine relativ schlechte Anpassungsgüte bezüglich der Punktwolke aufweist. Diesen unterschiedlichen Grad der Streuung um die Regressionsgerade wird durch das Größenverhältnis der Bestimmtheitsmaße $R^2_1 (a) > R^2_2 (b)$ in der Abbildung zum Ausdruck gebracht.

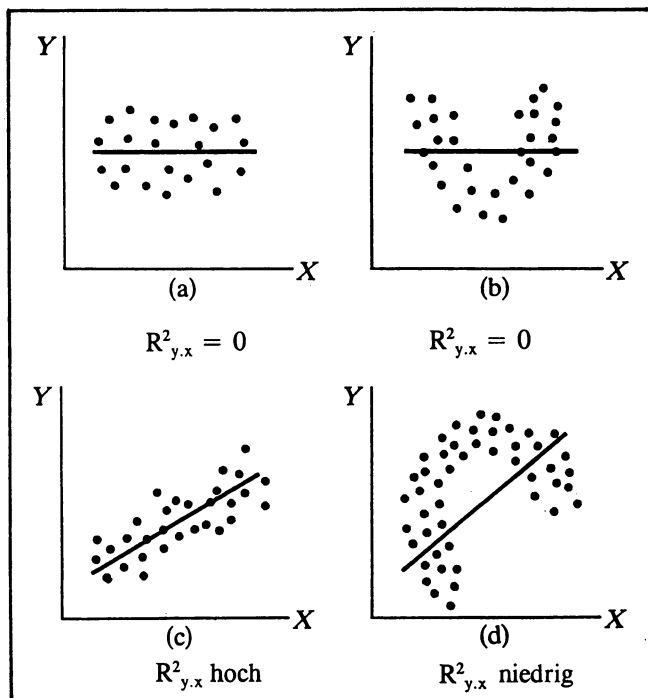


Abb. 5.14: Illustration unterschiedlich hoher Bestimmtheitsmaße

$$R^2_1(a) > R^2_2(b)$$

$$\begin{aligned} a_1 &= a_2 \\ b_1 &= b_2 \end{aligned}$$

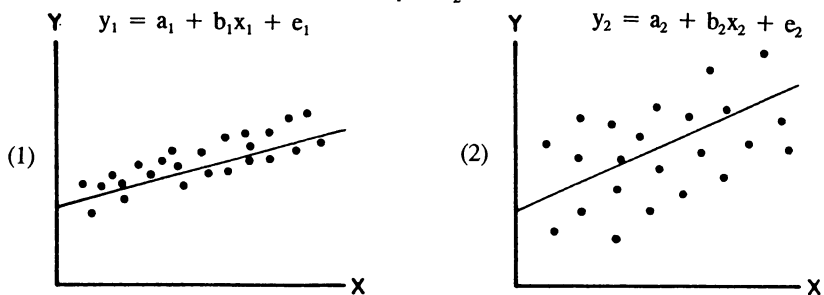


Abb. 5.15: Unterschiedliche Modellangepassungsgüten bei gleichen Regressionskoeffizienten

Quelle: Lewis-Beck, M.S., 1993: Applied regression: An introduction, in: Lewis-Beck, M.S. (ed), 1993: Regression analysis, International handbooks of quantitative applications in the social sciences, Vol. 2, Thousand Oaks, S. 17.

Analoge Betrachtungen können für die *nicht erklärte Varianz* angestellt werden. Je größer diese Varianz anteilmäßig ausfällt, desto unbestimmter ist die Abhängigkeit zwischen den Merkmalen Y und X. Folglich kann die nicht erklärte Varianz zur Charakterisierung des Unbestimmtheitsgrades einer Regression verwendet werden. Wird die Fehlervarianz durch die Gesamtvarianz dividiert, so erhält man das *Unbestimmtheitsmaß*:

$$1 - R^2 = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Der nicht erklärte Varianzanteil $1 - R^2_{y,x}$ (abgekürzt: $E^2_{y,x}$) wird auch als *Alienationskoeffizient* bezeichnet. Das Verhältnis der beiden Größen $R^2_{y,x}$ und $(1 - R^2_{y,x})$ läßt sich in einer Zerlegungsgleichung für die einzelnen Variationen angeben:

$$\frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} + \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$\frac{\text{Gesamtvariation}}{\text{Gesamtvariation}} = \frac{\text{Erklärte Variation}}{\text{Gesamtvariation}} + \frac{\text{Nicht erklärte Variation}}{\text{Gesamtvariation}}$$

$$1 = \text{Variationsanteil, der \underline{erklärt} ist} + \text{Variationsanteil, der \underline{nicht erklärt} ist}$$

$$1 = R^2 + 1 - R^2$$

Daraus läßt sich unmittelbar wieder die Varianzzerlegung des Regressionsmodells ableiten:

$$s_y^2 = R^2_{y,x} \cdot s_y^2 + (1 - R^2_{y,x}) \cdot s_y^2$$

$$s_y^2 = s_y^2 + s_e^2$$

Bevor wir uns einem Beispiel zuwenden, soll folgenden noch gezeigt werden, welche Beziehungen zwischen dem linearen Korrelationskoeffizienten r , dem Steigungskoeffizienten b der linearen Einfachregression und dem Bestimm-

heitsmaß $R^2_{y,x}$ bestehen. Vergleichen wir zunächst die Gleichung des Korrelationskoeffizienten mit der Gleichung des Steigungskoeffizienten der Einfachregression:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \cdot \frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x s_y}$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}$$

Der Korrelationskoeffizient läßt sich als Verhältnis der Kovarianz zweier Merkmale ($\text{cov}(x,y)$) zum Produkt der Standardabweichungen s_x und s_y darstellen, während der Steigungskoeffizient $b_{y,x}$ aus der Kovarianz (x, y) dividiert durch die Varianz des Merkmals X folgt. Vergleichen wir die beiden Gleichungen, so wird deutlich, daß der Korrelationskoeffizient r in den Steigungskoeffizienten $b_{y,x}$ überführbar ist, wenn die Gleichung für r mit s_y multipliziert und durch s_x dividiert wird:

$$\frac{\text{cov}(x, y)}{s_y s_x} \cdot \frac{s_y}{s_x} = \frac{\text{cov}(x, y)}{s_x^2} = b_{y,x}$$

Somit gilt:

$$b_{y,x} = r \cdot \frac{s_y}{s_x}$$

Der Regressionskoeffizient ist gleich dem Korrelationskoeffizient multipliziert mit dem Verhältnis der Standardabweichungen vom abhängigen und unabhängigen Merkmal. Damit läßt sich auch der Korrelationskoeffizient aus dem Steigungskoeffizienten berechnen:

$$r_{yx} = b_{yx} \cdot \frac{s_x}{s_y}$$

Die rechte Seite dieser Gleichung wird auch als *standardisierter Regressionskoeffizient* bezeichnet oder als Beta*-Koeffizient (symbolisch $\beta_{y,x}^*$). Der Beta*-Koeffizient ist in dem bivariaten Modell gleich dem Regressionskoeffizienten multipliziert mit dem Verhältnis der beiden Standardabweichungen s_x und s_y :

$$\beta_{y,x}^* = r_{yx} = b_{yx} \cdot \frac{s_x}{s_y}$$

Formal läßt sich ferner zeigen, daß der Korrelationskoeffizient sich als Quadratwurzel aus dem Produkt zweier Regressionskoeffizienten ergibt: In der einen Analyse wird Y als abhängiges Merkmal betrachtet und X als das unabhängige Merkmal, in der zweiten Analyse ist es genau umgekehrt, d.h. hier wird von X als abhängigem Merkmal ausgegangen:

$$b_{yx} = \frac{\text{cov}(x, y)}{s_x^2}$$

$$b_{xy} = \frac{\text{cov}(x, y)}{s_y^2}$$

$$r_{xy} = \sqrt{b_{xy} \cdot b_{yx}}$$

$$r_{xy} = \sqrt{\frac{\text{cov}(x, y)}{s_x^2} \cdot \frac{\text{cov}(x, y)}{s_y^2}} = \frac{\text{cov}(x, y)}{s_x s_y}$$

Es soll nun formal gezeigt werden, daß das Bestimmtheitsmaß der Regressionsanalyse mit dem quadrierten linearen Korrelationskoeffizienten nach Bravais-Pearson übereinstimmt. Den Korrelationskoeffizienten r hatten wir bereits als Zusammenhangsmaß für metrische Merkmale aus einer anderen Modellvorstellung abgeleitet. Die Quadrierung des Korrelationskoeffizienten ergibt:

$$r_{xy}^2 = \frac{\left[\sum (x_i - \bar{x}) (y_i - \bar{y}) \right]^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$$

Erweitert man den Zähler und den Nenner mit der Summe der Abweichungsquadrate von $(x_i - \bar{x})^2$, $i = 1, \dots, n$, und berücksichtigt, daß sich der Regres-

sionskoeffizient $b_{y,x}$ aus der Kovarianz (X, Y) dividiert durch die Standardabweichung von X ergibt, so folgt für den quadrierten Korrelationskoeffizienten:

$$r^2 = \frac{b^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (b x_i - b \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Da

$$\hat{y} = a_0 + b x_i \quad \text{und} \quad \bar{y} = a_0 + b \bar{x}$$

$$\bar{y} = a_0 + b x_i - a_0 - b \bar{x} = b x_i - b \bar{x}$$

gilt, läßt sich der quadrierte Korrelationskoeffizient wie folgt angeben:

$$r_{xy}^2 = \frac{\sum_{i=1}^n (b x_i - b \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\frac{1}{n} \sum_{i=1}^n (\hat{y} - \bar{y})^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{s_{\hat{y}}^2}{s_y^2}$$

Der Korrelationskoeffizient r ist somit gleich der Quadratwurzel aus dem Bestimmtheitsmaß der bivariaten Regressionsanalyse.

$$r_{xy}^2 = s_e^2 / s_y^2$$

$$\text{d.h. } r_{xy} = \sqrt{R_{y,x}^2}$$

Wir wollen abschließend die beschriebene Modelllogik der einfachen linearen Regressionsanalyse an einem bereits untersuchten Merkmalszusammenhang illustrieren. Wir gehen im folgenden wiederum von 422 Kreisen der fünf ausgewählten Regionen Köln-Aachen, Oberbayern-Schwaben, Niederbayern, Franken und Württemberg aus. Als abhängiges Merkmal betrachten wir den NSDAP-Stimmenanteil bei der Reichstagswahl von 1933, als unabhängiges Merkmal sei wiederum der Prozentanteil der Katholiken angenommen. Wenden wir die lineare Einfachregression auf die Ausgangsdaten an, erhalten wir die in der Tabelle 5.15 zusammengefaßten Ergebnisse.

Betrachten wir zunächst die Zahlenwerte für die Koeffizienten der Regressionsfunktion. Der Wert $b_{y,x} = -0,17$ besagt, daß mit der Zunahme des unabhängigen Merkmals X (Katholikenanteil) um einen Prozentpunkt der NSDAP-Stimmenanteil um 0,17% fällt. Bei einem Katholikenanteil von 0% erhalten wir

Tab. 5.15: NSDAP-Stimmenanteil und Katholikenanteil: Ergebnisse der einfachen Regressionsanalyse

Streuungs- quelle	Source of Variation	Summe der Abweichungs- quadrate (Sum of Squares)	Bestimmtheits- maß (Güte der Modellan- passung), Korrelations- koeffizient	Regressions- funktion
Erklärende Variable KATH%	Regression	17.212,42	$R^2_{y,x} = 0,308$ $r = - 0.555$	NSDAP1933 % = 46,5 - 0,173KATH%
Rest (Fehler)	Error	38.609,89	$1 - R^2_{y,x}$ = 0,692	---
Gesamt- variation	Total	55.822,31	---	---

als geschätzten NSDAP-Stimmenanteil einen Wert von 46,5% (gleich der Regressionskonstanten a). Ist die Regressionsgerade bestimmt, so kann der Vorhersagewert für den NSDAP-Stimmenanteil für jeden beobachteten Katholikenanteil berechnet werden. Wir erhalten z.B. für einen Katholikenanteil von 50% den Vorhersagewert 37,85% (= 46,5 – 0,173*50).

In der mittleren Spalte der Tabelle 5.15 sind die berechneten Abweichungsquadratsummen aufgenommen. Die entsprechenden Varianzen (bzw. Standardabweichungen) für die einzelnen Variationsquellen erhalten wir jeweils aus einer Division der einzelnen Werte durch n:

- Als Varianz von Y den Wert $55.822,31/422 = 132,28$ (Standardabweichung: 11,50);
- als Varianz der Störgröße den Wert $38.609,89/422 = 91,49$ (Standardabweichung: 9,56);
- als Varianz der mittels der linearen Regressionsfunktion vorhergesagten Y-Werte den Wert $17.212,42/422 = 40,78$ (Standardabweichung: 6,38).

Bei allen drei Variationsquellen ist die Summe der Abweichungen gleich Null und die Summe der quadrierten Abweichungen ein Minimum. Ein Vergleich der Varianz der Fehlerwerte (= 91,49) mit der Varianz der Y-Werte (= 132,28) führt zu der Feststellung, daß der Vorhersagefehler deutlich reduziert wird, wenn die Information der Katholikenanteile vermittelt der linearen Regressionsgleichung berücksichtigt wird (gleich dem aktuellen Modell im Gegensatz zum Basismodell, in dem lediglich der Mittelwert von Y zur Vorhersage der einzelnen NSDAP-Stimmenanteile herangezogen wird). Die Abweichungen der Regressionswerte auf der Geraden vom arithmetischen Mittel von Y weisen eine Varianz von 40,78 auf.

Bei der Berechnung der Güte der Modellanpassung kann direkt von den Abweichungsquadratsummen ausgegangen werden. Dividieren wir die erklärte Abweichungsquadratsumme (= 17.212,42) durch die Gesamtabweichungsquadratsumme (= 55.822,31) erhalten wir für das Bestimmtheitsmaß $R^2_{y,x}$ einen Wert von 0,308; dieser Wert entspricht einem Prozentanteil erklärter Varianz von 30,8 in dem abhängigen Merkmal »NSDAP-Stimmenanteil« nach Einführung des erklärenden Merkmals »Katholikenanteil« in das lineare Regressionsmodell. Aus dem Komplement des Bestimmtheitsmaßes folgt der nicht erklärte Varianzanteil mit einem Wert von 0,692 (d.h. 69,2% der Varianz in Y werden nicht durch den Katholikenanteil erklärt). Für den Korrelationskoeffizienten r als Zusammenhangsmaß zweier metrischer Merkmale erhalten wir einen Wert von -0,555. In dem bivariaten Modell ist r mit der Quadratwurzel aus $R^2_{y,x}$ identisch (d.h. die Quadratwurzel aus 0,308 ist gleich 0,555).

Im Rahmen der Regressions- und Korrelationsanalyse haben wir das metrische Skalenniveau der Merkmale »Katholikenanteil« und »NSDAP-Stimmenanteil« berücksichtigt. In der Kontingenztabellenanalyse hatten wir zur Veranschaulichung der Maßzahlen ordinaler Assoziation zunächst eine Klassifikation dieser Merkmale vorgenommen (den NSDAP-Stimmenanteil nach den Quartilen, den Katholikenanteil jeweils in Klassen der Breite von 33,33%). Der Wert für Kendall's Tau-c fällt mit -0,41 niedriger aus als der Korrelationskoeffizient r mit einem Wert von -0,55. Für den PRE-Koeffizienten GAMMA erhalten wir in der Tabellenanalyse einen Wert von 0,53; das PRE-Modell der Regressionsanalyse ergibt einen deutlich niedrigeren Determinationskoeffizienten von 0,308. Die unterschiedlichen Ergebnisse sind zum einen auf den erheblichen Informationsverlust aufgrund der Klassifikation der Ursprungswerte im Rahmen der Tabellenanalyse zurückzuführen. Zum anderen führen auch die unterschiedlichen Modellkonzepte zu divergierenden Werten. Als Regel läßt sich hier formulieren: Bei der Berechnung von Maßzahlen bivariater Assoziation sollte man immer das höchste Skalenniveau der Merkmale ausschöpfen.

Teil 3: Auswahlbibliographie

1. Forschungsmethodologie, Wissenschaftstheorie und historische Methode

- Abrahamson, M., 1983: Social research methods. Englewood Cliffs, NY: Prentice Hall.
- Acham, K., 1974: Analytische Geschichtsphilosophie. Freiburg: Alber.
- Adorno, T.W. u.a. (Hrsg.), 1982: Der Positivismusstreit in der deutschen Soziologie, 10. A. Neuwied: Luchterhand.
- Albert, H., 1991: Traktat über kritische Vernunft, 5. A. Tübingen: Mohr.
- Albert, H./Topitsch, E. (Hrsg.), 1971: Werturteilsstreit. Darmstadt: Wiss. Buchgesellschaft.
- Albert, H. (Hrsg.), 1972: Theorie und Realität. Tübingen: Mohr.
- Albert, H./Stapf, K.H. (Hrsg.), 1979: Theorie und Erfahrung. Stuttgart: Klett-Cotta.
- Alemann, H. von, 1977: Der Forschungsprozeß. Eine Einführung in die Praxis der empirischen Sozialforschung. Stuttgart: Teubner.
- Apel, K.-O./Tuomela, R. (Hrsg.), 1978: Neue Versuche über Erklären und Verstehen. Frankfurt a.M.: Suhrkamp.
- Atteslander, P., 1993: Methoden der empirischen Sozialforschung, 7. bearb. A. Berlin: de Gruyter.
- Babbie, E.R., 1989: The practice of social research, 5. ed. Belmont: Wadsworth.
- Bailey, K.D., 1987: Methods of social research, 3rd. ed. New York: Free Press.
- Baker, A.R.H./Billinge, M. (eds.), 1982: Period and place: Research methods in historical geography. Cambridge: Cambridge Univ. Press.
- Baker, T.L., 1988: Doing social research. New York: McGraw-Hill.
- Balzer, W., 1982: Empirische Theorien: Modelle, Strukturen, Beispiele. Braunschweig: Vieweg.
- Balzer, W./Moulines, C.U./Sneed, J.D., 1987: An architecture for science. The structuralist program. Dordrecht: Reidel.
- Baumgartner, H.M., 1972: Kontinuität und Geschichte. Frankfurt a.M.: Suhrkamp.
- Baumgartner, H.M./Rüsen, J. (Hrsg.), 1976: Seminar Geschichte und Theorie: Umriss einer Historik. Frankfurt a.M.: Suhrkamp.
- Berkhofer, R.F., 1969: A behavioral approach to historical analysis. New York: Wiley.
- Black, J.A., 1976: Methods and issues in social research. New York: Wiley.
- Bobinska, C., 1967: Historiker und historische Wahrheit. Zu erkenntnistheoretischen Problemen der Geschichtswissenschaft. Berlin.

- Bogue, A.G., 1983: *Clio and the bitch goddess: Quantifications in American political history*. Beverly Hills: Sage.
- Botz, G. (Hrsg.), 1988: »Qualität und Quantität«: Zur Praxis der Methoden der historischen Sozialwissenschaft. Frankfurt a.M.: Campus.
- Braithwaite, R.B., 1959: *Scientific explanation. A study of the function of theory, probability and law in science*. Cambridge: Univ. Press.
- Bryman, A., 1988: *Quantity and quality in social research*, 2nd improved ed. London: Unwin Hyman.
- Cahnman, W.J./Boskoff, A. (eds.), 1964: *Sociology and history: Theory and research*. New York: MacMillan.
- Chadwick, B./Bahr, H.M./ Albrecht, S.L., 1984: *Social science research methods*. Englewood Cliffs, NY: Prentice Hall.
- Cohen, L.J., (ed.), 1982: *Logic, methodology and philosophy of science*. Amsterdam: North-Holland.
- Cooper, H.M., 1989: *Integrating research*, 2nd ed. Beverly Hills: Sage.
- Crano, W.D./Brewer, M.B., 1986: *Principles and methods of social research*. Boston: Allyn & Bacon.
- Dale, A.S./Procter, M., 1988: *Doing secondary analysis*. London: Unwin Hyman.
- Danto, A.G., 1965: *Analytical philosophy of history*. Cambridge: Univ. Press.
- Denz, H., 1989: *Einführung in die empirische Sozialforschung*. Wien u.a.: Springer.
- Denzin, N., 1989: *The research act*, 3rd ed. Englewood Cliffs, NY: Prentice Hall.
- Dixon, B.R./Bouma, G.D./Atkinson, G.B.J., 1987: *A handbook of social science research*. Oxford: Oxford Univ. Press.
- Dooley, D., 1990: *Social research methods*, 2nd. ed. Englewood Cliffs, NY: Prentice-Hall.
- Dovring, F., 1960: *History as a social science. An essay on the nature and purpose of historical studies*. Den Haag.
- Dray, W.H., 1964: *Laws and explanation in history*, 3rd impr. repr. Oxford: Univ. Press.
- Droysen, J.G., 1960: *Historik: Vorlesungen über Enzyklopädie und Methodologie der Geschichte*, 4. A. Darmstadt: Luchterhand.
- Eberlein, G.L./Kroeber-Riel, W./Leinfellner, W. (Hrsg.), 1974: *Forschungslogik der Sozialwissenschaften*. Düsseldorf: Bertelsmann.
- Eckhardt, K./Ermann, M.D., 1977: *Social research methods*. New York: Random House.
- Engelbert, E. (Hrsg.), 1972: *Probleme der Geschichtsmethodologie*. Berlin: Springer.
- Esser, H./Klenovits, K./Zehnpfennig, H., 1977: *Wissenschaftstheorie*. 2 Bände. Stuttgart: Teubner.
- Essler, W.K., 1970–1979: *Wissenschaftstheorie*, 4 Bände. Freiburg: Alber.

- Faber, K.-G., 1972: *Theorie der Geschichtswissenschaft*, 2. durchges. A. München: Beck.
- Fain, H., 1970: History as science, in: *History and Theory*, 9, S. 154–173.
- Feyerabend, P.K., 1976: *Wider den Methodenzwang: Skizze einer anarchistischen Erkenntnistheorie*. Frankfurt a.M.: Suhrkamp.
- Fischer, D.H., 1970: *Historians' fallacies. Toward a logic of historical thought*. New York.
- Fischer, F.M., 1960: On the analysis of history and the interdependence of the social sciences, in: *Philosophy of Science*, 27, S. 147–158.
- Fogel, R.W./Elton, G.R., 1983: *Which road to the past? Two views of history*. New Haven: Yale Univ. Press.
- Forcese, D.P., 1973: *Social research methods*. Englewood Cliffs: Prentice-Hall.
- Frankfort-Nachmias, C./Nachmias, D., 1992: *Research methods in the social sciences*, 4. ed., reprint. London u.a.: Arnold.
- Friedrichs, J., 1990: *Methoden der empirischen Sozialforschung*, 14. A. Opladen: Westdt. Verlag.
- Gadamer, H.G., 1960: *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*. Tübingen: Mohr.
- Gallie, W., 1964: *Philosophy and the historical understanding*. London: Chatto & Windus.
- Galtung, J., 1967: *Theory and methods of social research*, 2nd impr. ed. London: Allen & Unwin.
- Galtung, J., 1978: *Methodologie und Ideologie*. Frankfurt a.M.: Suhrkamp.
- Gardiner, P., 1974: *The philosophy of history*. Oxford: Univ. Press.
- Geißler, R., 1973: *Soziologische Historie und historische Soziologie*, in: *Geschichte in Wissenschaft und Unterricht*, 24, S. 65–87.
- Giesen, B./Schmid, M. (Hrsg.), 1975: *Theorie, Handeln und Geschichte*. Hamburg: Hoffmann und Campe.
- Giesen, B./Schmid, M., 1976: *Erklärung und Geschichte*. Gersthofen: Maro.
- Goldenberg, S., 1992: *Thinking methodologically*. New York: Harper Collins.
- Greenberg, J./Folger, R., 1988: *Controversial issues in social research methods*. New York: Springer.
- Hakim, C., 1982: *Secondary analysis in social research. A guide to data sources and methods with examples*. London: Allen & Unwin.
- Hakim, C., 1987: *Research design: Strategies and choices in the design of social research*. London: Allen & Unwin.
- Hartman, J.J./Hedblom, J.H., 1979: *Methods for the social sciences*. Westport: Greenwood Press.
- Haskins, L./Jeffrey, K., 1990: *Understanding quantitative history*. Cambridge: MIT Press.
- Hempel, C.G., 1974: *Grundzüge der Begriffsbildung in der empirischen Wissenschaft*. Düsseldorf: Bertelsmann.
- Hempel, C.G., 1965: *Aspects of scientific explanation and other essays in the philosophy of science*. London: MacMillan.

- Hempel, C.G., 1977: Aspekte wissenschaftlicher Erklärung. Berlin/New York: de Gruyter.
- Hughes, H. S., 1960: The historian and the social scientist, in: *American Historical Review*, 66, S. 20–46.
- Hunt, M., 1991: Die Praxis der Sozialforschung. Reportagen aus dem Alltag einer Wissenschaft. Frankfurt a.M.: Campus.
- Hyman, H.H., 1972: *Secondary analysis of sample surveys: Principles, procedures and potentialities*. New York: Wiley.
- Iggers, G.G., 1971: *Deutsche Geschichtswissenschaft. Eine Kritik der traditionellen Geschichtswissenschaft von Herder bis zur Gegenwart*. München: Dt. Taschenbuchverlag.
- Iggers, G.G., 1978: *Neue Geschichtswissenschaft. Vom Historismus zur historischen Sozialwissenschaft*. München: Dt. Taschenbuchverlag.
- Iggers, G.G., 1993: *Geschichtswissenschaft im 20. Jahrhundert. Ein kritischer Überblick im internationalen Zusammenhang*. Göttingen: Vandenhoeck & Ruprecht.
- Imhof, A.E., 1977: *Einführung in die Historische Demographie*. München: Beck.
- Janich, P./Kambartel, F./Mittelstraß, J., 1974: *Wissenschaftstheorie als Wissenschaftskritik*. Frankfurt a.M.: Aspekte Verlag.
- Kaelble, H., 1978: *Historische Mobilitätsforschung. Westeuropa und die USA*. Darmstadt: Wiss. Buchgesellschaft.
- Kambartel, F. (Hrsg.), 1974: *Praktische Philosophie und konstruktive Wissenschaftstheorie*. Frankfurt a.M.: Suhrkamp.
- Kambartel, F., 1976: *Theorie und Begründung*. Frankfurt a.M.: Suhrkamp.
- Karrenberg, F./Albert, H. (Hrsg.), 1963: *Sozialwissenschaft und Gesellschaftsgestaltung*. Berlin: Duncker & Humboldt.
- Kelly, J.R./McGrath, J.E., 1988: *On time and method*. Beverly Hills: Sage.
- Kern, H., 1982: *Empirische Sozialforschung. Ursprünge, Ansätze, Entwicklungslinien*. München: Beck.
- Kocka, J. (Hrsg.), 1977: *Theorien in der Praxis des Historikers*. Göttingen: Vandenhoeck & Ruprecht.
- Kocka, J., 1986: *Sozialgeschichte. Begriff – Entwicklung – Probleme*, 2., erw. A. Göttingen: Vandenhoeck & Ruprecht.
- Kreppner, K., 1975: *Zur Problematik des Messens in den Sozialwissenschaften*. Stuttgart: Klett.
- Kromka, F., 1984: *Sozialwissenschaftliche Methodology*. Paderborn: Schöningh.
- Kromrey, H., 1991: *Empirische Sozialforschung. Modelle und Methoden der Datenerhebung und Datenauswertung*, 5., überarb. u. erw. A. Opladen: Leske + Budrich (UTB).
- Kuhn, T.S., 1967: *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt a.M.: Suhrkamp.

- Kuhn, T.S./Krüger, L. (Hrsg.), 1978: Die Entstehung des Neuen. Frankfurt a.M.: Suhrkamp.
- Kutschera, F. von, 1972: Wissenschaftstheorie. 2 Bände. München: Fink.
- Laatz, W., 1993: Empirische Methoden. Ein Lehrbuch für Sozialwissenschaftler. Frankfurt a.M.: Thun.
- Lakatos, I./Musgrave, A. (Hrsg.), 1974: Kritik und Erkenntnisfortschritt. Braunschweig: Vieweg.
- Landes, D.S./Tilly, C. (eds.), 1971: History as social science. Englewood Cliffs: Prentice-Hall.
- Lazarsfeld, P./Rosenberg, M. (eds.), 1955: The language of social research. Glencoe, Ill.: Free Press.
- Le Roy Ladurie, E., 1979: The territory of the historian. Stanford Terrace: Harvester Pr.
- Leinfellner, W., 1965: Struktur und Aufbau wissenschaftlicher Theorien. Wien: Physica.
- Leinfellner, W., 1967: Einführung in die Erkenntnis- und Wissenschaftstheorie, 2. erw. A. Mannheim: Bibliogr. Institut.
- Lenk, H., 1972: Erklärung, Prognose, Planung. Freiburg: Rombach.
- Leuzen, P., 1974: Konstruktive Wissenschaftstheorie. Frankfurt a.M.: Suhrkamp.
- Ludz, P.C. (Hrsg.), 1973: Soziologie und Sozialgeschichte. Opladen: Westdt. Verlag.
- Mayntz, R./Holm, K./Hübner, P., 1971: Einführung in die Methoden der empirischen Soziologie, 2., erw. A. Opladen: Westdt. Verlag.
- Meier, C./Rüsen, J. (Hrsg.), 1988: Historische Methode. München: Deutscher Taschenbuch Verlag.
- Miller, D.C., 1991: Handbook of research design and social measurement, 5. ed. London: Sage.
- Meran, J., 1985: Theorien in der Geschichtswissenschaft. Göttingen: Vandenhoeck & Ruprecht.
- Mittelstraß, J., 1974: Die Möglichkeit von Wissenschaft. Frankfurt a.M.: Suhrkamp.
- Mommsen, H., 1961: Historische Methode, in: Besson, W. (Hrsg.), 1961: Geschichte (= Fischer-Lexikon Bd. 24). Frankfurt a.M.: Fischer, S. 78–91.
- Neurath, O., 1979: Wissenschaftliche Weltauffassung, Sozialismus und Logischer Empirismus. Frankfurt a.M.: Suhrkamp.
- Nippert, R., 1972: Quantifizierung der sozialen Realität. Düsseldorf: Bertelsmann.
- Opp, K.-D., 1976: Methodologie der Sozialwissenschaften. Einführung in die Probleme ihrer Theoriebildung, durchges., rev. u. wesentl. erw. Neuauflage. Reinbek b. Hamburg: Rowohlt.
- Patzelt, W.J., 1986: Sozialwissenschaftliche Forschungslogik. München u.a.: Oldenbourg.

- Popper, K.R., 1965: Das Elend des Historizismus. Tübingen: Mohr.
- Popper, K.R., 1971: Logik der Forschung, 4., verb. A. Tübingen: Mohr.
- Postan, M., 1971: Fact and relevance. Essays on historical method. Cambridge: Univ. Press.
- Price, J.M./Nadel, G.H. (eds.), 1969: Studies in quantitative history and the logic of the social sciences. Middletown: Wesleyan Univ. Press.
- Prim, R./Tilman, H., 1989: Grundlagen einer kritisch-rationalen Sozialwissenschaft, 6., durchges. A. Stuttgart/Heidelberg: Quelle & Meyer (UTB).
- Rescher, N., 1970: Scientific explanation. New York: Free Press.
- Rossi, P., 1987: Vom Historismus zur historischen Sozialwissenschaft. Frankfurt a.M.: Suhrkamp.
- Roth, E., 1993: Sozialwissenschaftliche Methoden, 3., völlig überarb. u. erw. A. München u.a.: Oldenbourg.
- Rowney, D.K./Graham, J.Q. (eds.), 1969: Quantitative history: Selected readings in the quantitative analysis of historical data. Homewood: Dorsey Press.
- Rowney, D.K. (ed.), 1984: Soviet quantitative history. Beverly Hills: Sage.
- Rürup, R./Hausen, K. (Hrsg.), 1977: Historische Sozialwissenschaft. Beiträge zur Einführung in die Forschungspraxis. Göttingen: Vandenhoeck & Ruprecht.
- Rüsen J./Baumgartner, H.M. (Hrsg.), 1975: Historische Objektivität. Aufsätze zur Geschichtstheorie. Göttingen: Vandenhoeck & Ruprecht.
- Rüsen, J., 1983: Historische Vernunft. Grundzüge einer Historik I: Die Grundlagen der Geschichtswissenschaft. Göttingen: Vandenhoeck & Ruprecht.
- Rüsen, J., 1986: Rekonstruktion der Vergangenheit. Grundzüge einer Historik II: Die Prinzipien der historischen Forschung. Göttingen: Vandenhoeck & Ruprecht.
- Rüsen, J., 1989: Lebendige Geschichte. Grundzüge einer Historik III: Formen und Funktionen des historischen Wissens. Göttingen: Vandenhoeck & Ruprecht.
- Ruloff, D., 1984: Geschichtsforschung und Sozialwissenschaft. München: Oldenbourg.
- Schieder, T./Gräubig, K. (Hrsg.), 1977: Theorieprobleme der Geschichtswissenschaft. Darmstadt.
- Schieder, Th., 1965: Geschichte als Wissenschaft. Eine Einführung. München/Wien: Oldenbourg.
- Schieder, Th., 1971: Unterschiede zwischen historischer und sozialwissenschaftlicher Methode, in: Festschrift für H. Heimpel zum 70. Geburtstag, Bd. 1. Göttingen: Vandenhoeck & Ruprecht.
- Schnell, R./Hill, P.B./Esser, E., 1992: Methoden der empirischen Sozialforschung, 3., überarb. u. erw. A. München/Wien: Oldenbourg.
- Schulze, W., 1974: Soziologie und Geschichtswissenschaft. München: Fink.
- Schwemmer, O., 1976: Theorie der rationalen Erklärung. München: Beck.

- Seiffert, H., 1969–1973: Einführung in die Wissenschaftstheorie, 2 Bände. München: Beck.
- Simon, J.L./Selvin, H.C. (eds.), 1969: Basic research methods in social science: The art of empirical investigation, 5th print. New York: Random House.
- Singleton, R., 1988: Approaches to social research. New York: Oxford Univ. Press.
- Social research methodology bibliography, 1987 – 1993: Rotterdam: SRM-Documentation Centre, Erasmus Univ. (Vol. 1 – Vol. 15).
- Stewart, D.W., 1984: Secondary research. Beverly Hills: Sage.
- Stegmüller, W., 1983: Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie, Bd. I: Erklärung, Begründung, Kausalität, 2. A. Berlin: Springer.
- Stinchcombe, A.L., 1978: Theoretical methods in social history. New York: Academic Press.
- Tilly, C. 1981: As sociology meets history. New York: Academic Press.
- Thompson, B. (ed.), 1989: Advances in social science methodology: A research annual. Greenwich: JAI Press.
- Topitsch, E./Kraft, V. (Hrsg.), 1960: Probleme der Wissenschaftstheorie. Wien: Springer.
- Topitsch, E. (Hrsg.), 1972: Logik der Sozialwissenschaften, 8. A. Köln: Kiepenheuer & Witsch.
- Topolski, J. (ed.), 1990: Narration and explanation: Contributions to the methodology of the historical research. Amsterdam: Rodopi.
- Toulmin, S., 1969: Einführung in die Philosophie der Wissenschaft. Göttingen: Vandenhoeck & Ruprecht.
- Toulmin, S., 1972: Human understanding. Princeton: Univ. Press.
- Toulmin, S., 1981: Voraussicht und Verstehen. Frankfurt a.M.: Suhrkamp.
- Tschopp, A., 1990: Modellhaftes Denken in der Soziologie. Eine Untersuchung formaler Modelle in der empirischen Sozialforschung und in der soziologischen Theoriebildung. Frankfurt a.M.: P. Lang.
- Weber, M., 1973: Gesammelte Aufsätze zur Wissenschaftslehre, 4., erneut durchges. A. Tübingen: Mohr.
- Wehler, H.-U. (Hrsg.), 1972: Geschichte und Soziologie. Köln: Kiepenheuer & Witsch.
- Wehler, H.-U., 1973: Geschichte als Historische Sozialwissenschaft. Frankfurt a.M.: Suhrkamp.
- Wehler, H.-U., 1975: Modernisierungstheorie und Geschichte. Göttingen: Vandenhoeck & Ruprecht.
- Wehler, H.-U. (Hrsg.), 1976: Geschichte und Soziologie. Köln: Luchterhand.
- Wehler, H.-U., 1980: Historische Sozialwissenschaft und Geschichtsschreibung. Studien zu Aufgaben und Traditionen deutscher Geschichtswissenschaft. Göttingen: Vandenhoeck & Ruprecht.
- Wenturis, N./Van hove, W./Dreier, V., 1992: Methodologie der Sozialwissenschaften. Tübingen: Francke (UTB).

Willigan, J.D./Lynch, K.A., 1982: Sources and methods of historical demography. New York: Academic Press.

Wright, G.H. von, 1974: Erklären und Verstehen. Frankfurt a.M.: Athenaeum.

Schriftenreihe: Theorie der Geschichte (Beiträge zur Historik). München: Deutscher Taschenbuch Verlag, 1988:

Band 1: Kosellek, R./Mommsen, W.J./Rüsen, J. (Hrsg.): Objektivität und Parteilichkeit in der Geschichtswissenschaft.

Band 2: Faber, K.-G./Meier, C. (Hrsg.): Historische Prozesse.

Band 3: Kocka, J./Nipperdey, T. (Hrsg.): Theorie und Erzählung in der Geschichte.

Band 4: Kosellek, R./Lutz, H./Rüsen, J. (Hrsg.): Formen der Geschichtsschreibung.

Band 5: Meier, C./Rüsen, J. (Hrsg.): Historische Methode.

2. Statistik

Die Statistik-Literatur ist inzwischen zahllos. Bei der folgenden Literaturauswahl wurden anwendungsbezogene Darstellungen vorwiegend aus den Sozial- und Wirtschaftswissenschaften berücksichtigt. Die Auswahlbibliographie erhebt selbstredend keinen Anspruch auf Vollständigkeit oder auf eine systematische Erfassung von Literaturhinweisen. Sie dokumentiert die Standardliteratur und umfaßt eine Aufschlüsselung in den jeweiligen Themenschwerpunkten nach weiterführenden, z.T. sehr spezifischen Fragestellungen (mit (S) gekennzeichnet; größtenteils mit hohem mathematischen Niveau).

Nicht berücksichtigt wurde die Spezialliteratur zur Auswertung experimenteller Versuchsanordnungen (sog. faktorielle Designs). Spezielle Monographien zu Fragestellungen der Varianz- und Kovarianzanalyse sind daher nicht gesondert aufgenommen worden. Diese Thematik läßt sich aber formalstatistisch als Spezialfall des allgemeinen linearen Modells ableiten. Daher wird die Literatur gesondert ausgewiesen, die diese Verfahren als zwei Varianten der multiplen Regression darstellen.

Die zahlreichen Beiträge aus den großen Sammlungen »Techniken der empirischen Sozialforschung«, »Die Befragung«, »Handbuch der empirischen Sozialforschung« und »Quantitative Applications in the Social Sciences« sind am Ende der Statistikbibliographie jeweils gesondert unter dem Gliederungspunkt »Sammlungen« berücksichtigt.

2.1 Auswahlverfahren

Bausch, T., 1990: Stichprobenverfahren in der Marktforschung. München: Vahlen.

- Böltkén, F., 1976: Auswahlverfahren. Stuttgart: Teubner.
- Chaudhuri, A./Vos, J.W.E., 1988: Unified theory and strategies of survey sampling. Amsterdam: North-Holland. (S)
- Cochran, W.G., 1972: Stichprobenverfahren. Berlin: de Gruyter.
- Cochran, W.G., 1977: Sampling techniques, 3rd. ed. New York: Wiley.
- Henry, G.T., 1990: Practical sampling. Newbury Park: Sage.
- Kellerer, H., 1963: Theorie und Technik des Stichprobenverfahrens, 3. A. München: Einzelschrift der Deutschen Statistischen Gesellschaft.
- Kish, L., 1965: Survey sampling. New York: Wiley.
- Kreienbrock, L., 1993: Einführung in die Stichprobenverfahren: Lehr- und Übungsbuch der angewandten Statistik, 2. durchges. A. München: Oldenbourg.
- Krewski, D./Platek, R., Rao, J.N.K., 1981: Current topics in survey sampling. New York: Academic Press.
- Levy, P.S., 1991: Sampling of populations: Methods and applications. New York u.a.: Wiley. (S)
- Murthy, M.N., 1967: Sampling theory and methods. Calcutta: Statistical Publ. Society. (S)
- Neubäumer, R., 1982: Die Eigenschaften verschiedener Stichprobenverfahren bei wirtschafts- und sozialwissenschaftlichen Untersuchungen. Frankfurt a.M.: Peter Lang. (S)
- Persell, C.H./Maisel, R., 1995: How sampling works. London: Sage.
- Scheaffer, R.L./Mendenhall, W./Ott, L., 1979: Elementary survey sampling. North Scituate, Mass.: Duxbury Press.
- Schwarz, H., 1975: Stichprobenverfahren: Ein Leitfaden zur Anwendung statistischer Schätzverfahren. München: Oldenbourg.
- Stenger, H. (Hrsg.), 1980: Praktische Anwendungen von Stichprobenverfahren. Göttingen. Vandenhoeck & Ruprecht.
- Stenger, H., 1986: Stichproben. Heidelberg: Physica.
- Stuart, A., 1984: The ideas of sampling. High Wycombe: Griffin.
- Sudman, S., 1976: Applied sampling. New York: Academic Press.
- Wauschkuhn, U., 1982: Anpassung von Stichproben und n-dimensionalen Tabellen an Randbedingungen. München/Wien: Oldenbourg.

2.2 Deskriptiv- und Inferenzstatistik

- Anderson, A.J., 1989: Interpreting data: a first course in statistics. London: Chapman & Hall.
- Andrews, F. u.a., 1981: A guide for selecting statistical techniques for analyzing social science data. Ann Arbor: Univ. of Michigan.
- Arney, W.R., 1990: Understanding statistics in the social sciences. New York: Freeman.
- Atteslander, P. u.a., 1991: Methoden der empirischen Sozialforschung, Berlin/New York: de Gruyter.

- Bakeman, R., 1992: Understanding social science data: a spreadsheet approach. Hillsdale, NY: Erlbaum.
- Bamberg, G./Baur, F., 1993: Statistik, 8., überarb. u. erw. A. München/Wien: Oldenbourg.
- Benninghaus, H., 1991: Einführung in die sozialwissenschaftliche Datenanalyse, 2., völlig überarb. A. München/Wien: Oldenbourg.
- Bley Müller, J./Gehlert, G./Gülicher, H., 1992: Statistik für Wirtschaftswissenschaftler, 8., überarb. A. München: Vahlen.
- Bohley, P., 1992: Statistik, 5., überarb. u. erg. A. München/Wien: Oldenbourg.
- Bol, G., 1992: Wahrscheinlichkeitstheorie: Einführung. München/Wien: Oldenbourg.
- Bol, G. 1993: Deskriptive Statistik: Einführung, 2., überarb. u. erw. A. München/Wien: Oldenbourg.
- Bomdsdorf, E., 1992: Induktive Statistik, 5., durchges. A. Köln: Verlag Josef Eul.
- Bortz, J., 1989: Lehrbuch der Statistik für Sozialwissenschaftler, 3. A., Berlin/Heidelberg.
- Bosch, K., 1991: Statistik für Nichtstatistiker. München/Wien: Oldenbourg.
- Bredenkamp, J., 1972: Der Signifikanztest in der psychologischen Forschung. Frankfurt a.M.: Akademische Verlagsgesellschaft.
- Brockett, P./Levine, A., 1984: Statistics and probability and their applications. Philadelphia/New York u.a.: Saunders College Publishing.
- Byrkit, D.R., 1987: Statistics today. A comprehensive introduction. Menlo Park u.a.: Benjamin/Cummings.
- Christensen, B., 1991: Introduction to statistics for the social and behavioral sciences. Pacific Grove: Brooks/Cole.
- Cox, C.P., 1987: Handbook of introductory statistical methods. New York: Wiley.
- DeGroot, M.H., 1989: Probability and statistics, 2nd ed., Reading, Mass. u.a.
- Devore, J./Peck, R., 1986: Statistics. The exploration and analysis of data. St. Paul u.a.: West Publishing Company.
- Dreier, V., 1994: Datenanalyse für Sozialwissenschaftler. München/Wien: Oldenbourg.
- Eckey, H.-F. u.a., 1992: Statistik: Grundlagen, Methoden, Beispiele. Wiesbaden: Gabler.
- Ehrenberg, A.S.C., 1986: Statistik oder der Umgang mit Daten. Eine praktische Einführung mit Übungen. Weinheim u.a.: VCH-Verlag.
- Elpelt, B./Hartung, J., 1992: Grundkurs Statistik. München/Wien: Oldenbourg.
- Erickson, B.H./Nosanchuk, T.A., 1992: Understanding data, 2nd ed. Buckingham: Open Univ. Press.
- Graff, J., 1992: Soziologische Statistik. Pfaffenweiler: Centaurus.
- Grimm, J.W., 1990: Basic social statistics and quantitative research methods: A computerassisted introduction. Belmont: Wadsworth.

- Härter, E., 1987: Wahrscheinlichkeitsrechnung – statistische und mathematische Grundlagen. Göttingen.
- Hafner, R., 1992: Statistik für Sozial- und Wirtschaftswissenschaftler. Wien u.a.: Springer.
- Hamilton, L.C., 1990: Modern data analysis. A first course in applied statistics. Pacific Grove: Brooks/Cole.
- Heiler, S./Michels, P., 1993: Deskriptive und Explorative Datenanalyse. München/Wien: Oldenbourg.
- Hellmund, U., 1992: Grundlagen der Statistik. Landsberg/Lech: Moderne Industrie.
- Hellevik, O., 1984: Introduction to causal analysis: Exploring data by cross-tabulation. London: Allen & Unwin.
- Holland, H./Scharnbacher, K., 1990: Grundlagen der Statistik. Datenerfassung und -aufbereitung, Darstellung des statistischen Materials, statistische Maßzahlen, Verhältnis- und Indexzahlen, Zeitreihenanalyse. Wiesbaden: Gabler.
- Hugill, M., 1985: Advanced statistics. A comprehensive A-level course. London: Bell & Hyman.
- Kanji, G.K., 1993: 100 statistical tests. Newbury Park: Sage.
- Kapadia, R./Andersson, G., 1987: Statistics explained – Basic concepts and methods. New York: Halsted Press.
- Kennedy, G., 1985: Einladung zur Statistik. Frankfurt a.M.: Campus.
- Krämer, W., 1991: So lügt man mit Statistik. New York: Campus.
- Kriz, J., 1973: Statistik in den Sozialwissenschaften. Reinbek b. Hamburg: Rowohlt.
- Kriz, J., 1981: Methodenkritik empirischer Sozialforschung. Stuttgart: Teubner.
- Laatz, W., 1993: Empirische Methoden: Ein Lehrbuch für Sozialwissenschaftler. Thun u.a.: Deutsch.
- Liebertson, S., 1985: Making it count. The improvement of social research and theory. Berkeley: Univ. of California Press.
- Lippe, P. von der, 1993: Deskriptive Statistik. Stuttgart/Jena: Fischer (UTB).
- Marascuilo, L.A./Serlin, R.C., 1988: Statistical methods for the social and behavioral sciences. New York: Freeman.
- Marsh, C., 1988: Exploring data. An introduction to data analysis for social scientists. Cambridge: Polity Press.
- Menges, G., 1982: Die Statistik: 12 Stationen des statistischen Arbeitens. Wiesbaden: Gabler.
- Milton, J.S./Corbet, J.J./McTeer, P.M., 1986: Introduction to statistics. Lexington, Mass.: D.C. Heath and Company.
- Moore, D.S./McCabe, G.P., 1993: Introduction to the practice of statistics, 2nd ed. New York: W.H. Freeman and Company.
- Moser, L.E., 1986: Think and explain with statistics. Reading, Mass.: Addison-Wesley.
- Neter, J., 1988: Applied statistics. Boston: Allyn & Bacon.

- Oakes, M., 1986: Statistical inference. A commentary for the social und behavioral sciences. Chichester: Wiley.
- Ott, L., 1987: Statistics. A tool for social scientists, 4th rev. ed. Boston: Duxbury. Ott, L., 1988: An introduction to statistical methods and data analysis, 3rd ed. Boston: PWS-Kent.
- Patzelt, W.J., 1985: Einführung in die sozialwissenschaftliche Statistik. München/Wien: Oldenbourg.
- Pfanzagl, J., 1967: Allgemeine Methodenlehre der Statistik, Bd. I. Berlin/New York: de Gruyter.
- Pinnekamp, H.-J./Siegmann, F., 1992: Deskriptive Statistik. Einführung in die statistische Methodenlehre, 2., erw. A. München/Wien: Oldenbourg.
- Pinnekamp, H.-J./Siegmann, F., 1993: Deskriptive Statistik. Mit einer Einführung in das Programmpaket STATGRAPHICS. 3. erw. u. bearb. A. München/Wien: Oldenbourg.
- Raghavarao, D., 1988: Exploring statistics. New York/Basel: Dekker.
- Randow, G. v., 1992: Das Ziegenproblem: Denken in Wahrscheinlichkeiten. Hamburg: Rowohlt.
- Reid, S., 1987: Working with statistics: An introduction to quantitative methods for social scientists. Cambridge: Polity Press.
- Rüger, B., 1988: Induktive Statistik, 2., überarb. A. München/Wien: Oldenbourg.
- Sachs, L., 1992: Angewandte Statistik: Anwendung statistischer Methoden, 7., völlig Neubearb. A. Berlin u.a.: Springer.
- Schaich, E., 1990: Schätz- und Testmethoden, 2., verb. A. München: Vahlen.
- Schlittgen, R., 1993: Einführung in die Statistik: Analyse und Modellierung von Daten, 4., überarb. und erw. A. München/Wien: Oldenbourg.
- Schneider, W./Kornumpf, J./Mohr, W., 1993: Statistische Methodenlehre. Definitions- und Formelsammlung zur deskriptiven und induktiven Statistik mit Erläuterungen. München: Oldenbourg.
- Schulze, P., 1994: Beschreibende Statistik, 2., überarb. u. erg. A. München/Wien: Oldenbourg.
- Siegel, A.F., 1988: Statistics and data analysis. An introduction. New York u.a.: Wiley
- Streck, B.H., 1991: Grundlagen der Statistik: eine praxisorientierte Einführung mit Fallbeispielen. Stuttgart: Kohlhammer.
- Urban, K., 1992: Statistik. Einführung in die statistische Methodenlehre für Wirtschafts- und Sozialwissenschaftler, 3., verb. u. aktualisierte A. München/Wien: Oldenbourg.
- Vogel, F., 1991: Beschreibende und schließende Statistik, 6. A. München u.a.: Oldenbourg.
- Vogt, W.P., 1993: Dictionary of statistics and methodology. Newbury Park: Sage.
- Voß, W., 1979: Bausteine der Statistik. Köln: Kiepenheuer.

- Wagenführ, R., 1971: Statistik leicht gemacht. Band I: Einführung in die deskriptive Statistik. Köln: Kiepenheuer.
- Walsh, A., 1990: Statistics for the social sciences. New York: Harper & Row.
- Wehrt, K., 1984: Beschreibende Statistik: Eine Einführung. Frankfurt a.M./New York: Campus.
- Wonnacott, T.H./Wonnacott, R.J., 1990: Introductory statistics, 5th ed. New York: Wiley.
- Zeisel, H., 1970: Die Sprache der Zahlen. Köln: Kiepenheuer.
- Zöfel, P., 1992: Statistik in der Praxis. Stuttgart: G. Fischer (UTB).

Lehrbücher für Historiker

- Clubb, J.M. et al. (eds.), 1989: The process of historical inquiry: Everyday lives of working Americans. New York.
- Dollar, C.M./Jensen, R.J., 1971: Historian's guide to statistics: Quantitative analysis and historical research. New York.
- Drake, M., 1974: The quantitative analysis of historical data. Milton Keynes.
- Floud, R., 1980: Einführung in quantitative Methoden für Historiker, Stuttgart. (deutsche Übersetzung der 2. A. von: An introduction to quantitative methods for Historians. London 1979).
- Haskins, L./Jeffrey, K., 1990: Understanding quantitative history. Cambridge, Mass.
- Heffer, J./Robert, J.-L./Saly, P., 1981: Outils statistiques pour les historiens. Paris.
- Jarusch, K.H./Arminger, G./Thaller, M., 1985: Quantitative Methoden in der Geschichtswissenschaft: Eine Einführung in die Forschung, Datenverarbeitung und Statistik. Darmstadt.
- Jarusch, K.H./Hardy, K.A., 1991: Quantitative methods for historians: A guide to research, data, and statistics. London.
- Loether, H./McTavish, D.G., 1988: Descriptive and inferential statistics: An introduction, 3rd rev. ed. Boston.
- Ohler, N., 1980: Quantitative Methoden für Historiker. München.

2.3 Explorative Datenanalyse

- Enke, H. (Hrsg.), 1992: Methoden und Werkzeuge für die explorative Datenanalyse in den Biowissenschaften. Stuttgart: G. Fischer.
- Hoaglin, D.C./Mosteller, F./Tukey, J.W. (eds.), 1983: Understanding robust and exploratory data analysis. New York: Wiley.
- Hoaglin, D.C./Mosteller, F./Tukey, J.W., 1985: Exploring data tables, trends and shapes. New York: Wiley.
- Hoaglin, D.C., 1991: Fundamentals of exploratory analysis of variance. New York: Wiley. (S)

- Lehmacher, V.W./van Eimeren, W. (Hrsg.), 1980: Explorative Datenanalyse. Berlin u.a.: Springer.
- Rutsch, M., 1988: Statistik 1 – Mit Daten umgehen, 2., überarb. u. erg. A. Basel u.a.: Birkhäuser.
- Polasek, W., 1988: EDA, explorative Datenanalyse. Berlin: Springer.
- Tukey, J.W., 1977: Exploratory data analysis. Reading, Mass.: Addison-Wesley.
- Velleman, P.F./Hoaglin, D.C., 1981: Applications, basics, and computing of exploratory data analysis. Boston: Duxbury.

2.4 Regressionsanalyse

- Barnett, V./Lewis, T., 1984: Outliers in statistical data, 2nd ed. New York: Wiley.
- Bates, D.M./Watts, D.G., 1988: Nonlinear regression analysis. New York: Wiley. (S)
- Belsley, D.A., 1991: Conditioning diagnostics: Collinearity and weak data in regression. New York: Wiley. (S)
- Carroll, R.J./Ruppert, D., 1988: Transformation and weighting in regression. New York: Chapman & Hall. (S)
- Chatterjee, S., 1977: Regression analysis by example. New York: Wiley.
- Chatterjee, S./Hadi, A.S., 1988: Sensitivity analysis in linear regression. New York: Wiley. (S)
- Cook, R.D./Weisberg, S., 1982: Residuals and influence in regression. New York: Wiley. (S)
- Dielman, T.E., 1990: Applied regression analysis. Boston: PWS-Kent.
- Doran, H.E., 1989: Applied regression analysis in econometrics. New York: Dekker. (S)
- Draper, N.R./Smith, H., 1981: Applied regression analysis, 2nd ed. New York: Wiley.
- Dunteman, G.H., 1984: Introduction to linear models. Beverly Hills: Sage.
- Edwards, A.L., 1985: An introduction to linear regression and correlation. New York: Freeman.
- Fox, J., 1991: Regression diagnostics. Newbury Park, Calif.: Sage. (S)
- Freund, R.J./Minton, P.D., 1979: Regression methods: A tool for data analysis. New York: Dekker.
- Gruber, M.H.J., 1990: Regression estimators: A comparative study. Boston: Academic Press. (S)
- Gunst, R.F./Mason, R.C., 1980: Regression analysis and its application. New York: Dekker.
- Hamilton, L.C., 1992: Regression with graphics. Belmont, Calif.: Wadsworth.
- Jaccard, J., 1990: Interaction effects in multiple regression. Newbury Park, Calif.: Sage. (S)
- Kockläuner, G., 1988: Angewandte Regressionsanalyse mit SPSS. Braunschweig: Vieweg.

- Miller, A.J., 1990: Subset selection in regression. London u.a.: Chapman & Hall. (S)
- Montgomery, D.C./Peck, E.A., 1982: Introduction to linear regression analysis. New York: Wiley.
- Myers, R.H., 1990: Classical and modern regression with applications. Boston: PWS-Kent.
- Neter, J./Wassermann, W./Kutner, M.H., 1988: Applied linear regression models, 2nd ed. Homewood: Irwin.
- Pokropp, F., 1994: Lineare Regression und Varianzanalyse. München/Wien: Oldenbourg.
- Ratkowsky, D.A., 1983: Nonlinear regression modelling: A unified practical approach. New York: Dekker. (S)
- Rawling, J.O., 1988: Applied regression analysis: A research tool. Pacific Grove: Wadsworth.
- Rönz, B./Förster, E., 1992: Regressions- und Korrelationsanalyse. Grundlagen -Methoden – Beispiele. Wiesbaden: Gabler. (S)
- Seber, G.A.F./Wild, C.J., 1989: Nonlinear regression. New York: Wiley. (S)
- Sen, A., 1993: Regression analysis. Theory, methods, and applications. 2nd corr. ed. Berlin: Springer. (S)
- Tiede, M., 1987: Statistik: Regressions- und Korrelationsanalyse. München: Oldenbourg.
- Urban, D., 1982: Regressionstheorie und Regressionstechnik. Stuttgart: Teubner.
- Wetherill, G.B., 1986: Regression analysis with applications. London: Chapman & Hall.
- Weisberg, S., 1985: Applied linear regression, 2nd ed. New York: Wiley.
- Wittink, D.R., 1988: The application of regression analysis. London: Allyn & Bacon.
- Younger, M.S., 1979: Handbook for linear regression. North Scituate, Mass.: Duxbury Press.

2.5 Regressions- und Varianzanalyse auf der Basis des Allgemeinen Linearen Modells

- Cohen, J./Cohen, P., 1983: Applied multiple regression/correlation analysis for the behavioral sciences, 2nd ed. Hilldale: Lawrence Erlbaum Associates.
- Dunn, O.J./Clark, V.A., 1987: Applied statistics: Analysis of variance and regression, 2nd ed. New York: Wiley.
- Edwards., A.L., 1979: Multiple regression and the analysis of variance and covariance. San Francisco: Freeman.
- Kerlinger, F.N./Pedhazur, E.J.: 1973: Multiple regression in behavioral research. New York u.a.: Holt.
- MacNeil, K.A./Kelly, F.J., 1975: Testing research hypotheses using multiple linear regression. Carbondale, Ill.: Southern Illinois Univ. Press.

- Moosbrugger, H./Klutky, N., 1987: Regressions- und Varianzanalysen auf der Basis des Allgemeinen Linearen Modells. Bern: Huber.
- Namboodiri, N.K., 1975: Applied multivariate analysis and experimental designs. New York. (S)
- Neter, J./Wassermann, W., 1974: Applied linear statistical modelling. Homewood, Ill.: Irwin.
- Rochel, H., 1983: Planung und Auswertung von Untersuchungen im Rahmen des allgemeinen linearen Modells. Berlin u.a.: Springer.
- Wesolowsky, G.O., 1976: Multiple regression and analysis of variance. New York: Wiley.

2.6 Fehlende Werte in der Statistik

- Jänner, M., 1993: Neue Ansätze zur Lösung des Problems fehlender Werte im linearen Regressionsmodell. Frankfurt a.M.: P. Lang. (S)
- Kalton, G., 1983: Missing survey data. Ann Arbor.
- Little, R.J.A./Rubin, D.B., 1987: Statistical analysis with missing data. New York: Wiley. (S)
- Madow, W.G./Nisselson, H./Olkin, I. (eds.), 1983: Incomplete data in sample surveys. Vol. 2: Theory and bibliographies. New York: Academic Press. (S)
- Schwab, G., 1991: Fehlende Werte in der angewandten Statistik. Wiesbaden: Dt. Universitäts-Verlag. (S)

2.7 Multivariate Datenanalysemethoden: Übersichten

- Afifi, A.A./Clark, V., 1990: Computer-aided multivariate analysis, 2nd ed. New York: Van Nostrand Reinhold.
- Amick, D.J./Walberg, H.J., (eds.), 1975: Introductory multivariate analysis. Berkely, Calif.: McCutchan.
- Backhaus, K. u.a., 1993: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 7., vollständig überarb. u. erw. A. Berlin u.a.: Springer.
- Barnett, V., 1981: Interpreting multivariate data. Chichester: Wiley.
- Bennett, S./Bowers, D., 1976: An introduction to multivariate techniques for social and behavioral sciences. New York: Wiley.
- Bernstein, I.H./Garbin, C.P./Teng, G.K., 1988: Applied multivariate analysis. New York: Springer.
- Berry, W.D./Lewis-Beck, M.S. (eds.), 1986: New tools for social scientists: Advances and applications in research methods. London: Sage.
- Bortz, J., 1989: Statistik für Sozialwissenschaftler, 3., neubearb. A. Berlin u.a.: Springer.
- Bryman, A./Cramer, O., 1990: Quantitative data analysis for social scientists. London: Routledge.

- Chatfield, C./Collins, A.J., 1980: Introduction to multivariate data analysis. London: Chapman & Hall.
- Dillon, W.R./Goldstein, M., 1984: Multivariate analysis: Methods and applications. New York: Wiley.
- Dunteman, G.H., 1984: Introduction to multivariate analysis. Beverly Hills: Sage.
- Everitt, B., 1983: Advanced methods of data exploration and modeling. London u.a.: Heinemann.
- Fahrmeir, L./Hamerle, A. (Hrsg.), 1984: Multivariate statistische Verfahren. Berlin/New York: Walter de Gruyter. (S)
- Flury, B./Riedwyl, H., 1983: Angewandte multivariate Statistik. Stuttgart/New York: G. Fischer.
- Fox, J., 1984: Linear statistical models and related methods. New York: Wiley. (S)
- Fox, J./Long, S. (eds.), 1990: Modern methods of data analysis. Newbury Park: Sage. (S)
- Green, P.E./Carroll, J.D., 1976: Mathematical tools for applied multivariate analysis. New York: Academic Press. (S)
- Hartung, J./Elpelt, B., 1992: Multivariate Statistik. München: Oldenbourg. (S)
- Hair, J.F./Anderson, R.E./Tatham, R.C., 1987: Multivariate data analysis with readings, 2nd. ed. New York/London: Macmillan. (S)
- Hanushek, E.A./Jackson, J.E., 1977: Statistical methods for social scientists. New York: Academic Press.
- Hawkins, D.M., 1982: Topics in applied multivariate analysis. Cambridge: Cambridge Univ. Press.
- Jambu, M., 1991: Exploratory and multivariate data analysis. Boston u.a.: Academic Press.
- Jobson, J. D., 1991: Applied multivariate data analysis. Volume I: Regression and Experimental Design. New York u.a.: Springer.
- Jobson, J. D., 1991: Applied multivariate data analysis. Volume II: Categorical and Multivariate Methods. New York u.a.: Springer.
- Johnson, R., 1992: Applied multivariate statistical analysis. New York: Prentice Hall.
- Judd, C.M., 1989: Data-analysis: A model-comparison approach. San Diego: Harcourt. (S)
- Karson, M.J., 1982: Multivariate statistical methods: An introduction. Ames: The Iowa State Univ. Press.
- Kendall, M.G./Stuart, A., 1966: The advanced theory of statistics. Vol. 1 – Vol. 3, 2nd ed. London: Griffin.
- Kleinbaum, D.G./Kupper, L.L./Muller, K.E., 1988: Applied multiple regression and other multivariate methods. Boston: PWS-Kent.
- Läuter, J., 1992: Stabile multivariate Verfahren: Diskriminanzanalyse – Regressionsanalyse – Faktorenanalyse. Berlin: Akademie Verlag.
- Lunn, D., 1987: Excursions in data analysis. Milton Keynes: Open Univ. Press.

- Mackay, D.H. (ed.), 1983: Data analysis in the social sciences. London:
- Manly, B.F.J., 1986: Multivariate statistical methods: A primer. London: Chapman & Hall.
- Marascuilo, L.A., 1983: Multivariate statistics in the social sciences. A researcher's guide. Monterey, Calif.: Brooks/Cole.
- Marinell, G., 1990: Multivariate Verfahren, 3., erw. A. München: Oldenbourg.
- Morrison, D.F., 1976: Multivariate statistical methods, 2nd ed. New York: McGraw-Hill.
- Moosbrugger, H., 1978: Multivariate statistische Verfahren. Stuttgart u.a.: Kohlhammer.
- Press, S.J., 1972: Applied multivariate analysis. New York: Holt.
- Schofield, N., 1986: Advanced statistical methods in the social sciences. New York: Praeger.
- Stevens, J., 1986: Applied multivariate statistics for the social sciences. Hillsdale, NY: Erlbaum.
- Tabachnick, B.G./Fidell, Linda S., 1983: Using multivariate statistics. New York: Harper & Row.
- Tatsuoka, M.M., 1988: Multivariate analysis, 2nd ed. New York/London: Macmillan.
- Van de Geer, J.P., 1971: Introduction to multivariate analysis for the social sciences. San Francisco: W.H. Freeman and Company.
- Wall, F.J., 1985: Statistical data analysis handbook. New York: McGraw-Hill.

2.8 Mehrebenenanalyse

- Alpheus, H., 1988: Kontextanalyse. Wiesbaden: Deutscher Universitäts-Verlag.
- Bock, R.D. (ed.), 1988: Multilevel analysis of educational data. New York: Academic Press. (S)
- Boyd, L.A./Iversen, G.R.: Contextual analysis: Concepts and statistical techniques. Belmont, Calif.: Wadsworth.
- Bryk, A.S. u.a., 1992: Hierarchical linear models: Application and data analysis methods. Newbury Park: Sage. (S)
- Clar, M.R., 1981: Methodologische Probleme der Mehrebenenanalyse. Hamburg: Selbstverlag. (Phil. Diss. Univ. Hamburg)
- Dar, Y./Resh, N. (eds.), 1986: Classroom composition and pupils' achievement. London: Gordon and Breach.
- Dogan, M./Rokkan, S. (eds.): Quantitative ecological analysis in the social sciences. Cambridge: Univ. Press.
- Eeden, P. van den/Hüttner, H.J.M., 1982: Trend-Report multilevel research, Beverly Hills.
- Eeden, P. van den/Hox, J./Hauer, J. (eds.), 1990: Theories and models in multilevel research: Convergence or divergence? Amsterdam: SISWO.
- Esser, H., 1988: Sozialökologische Stadtforschung und Mehr-Ebenen-Analyse, in: Friedrichs, J. (Hrsg.), 1988: Stadtsoziologie, Sonderheft 29 der Kölner Zeitschrift für Soziologie und Sozialpsychologie.

- Goldstein, H., 1987: Multilevel models in educational and social research. London: Oxford Univ. Press.
- Hannan, M.T., 1971: Aggregation and disaggregation in sociology. Lexington, Mass.: Heath.
- Hox, J. J./Kreft, I.G.G. (eds.), 1994: Multilevel analysis methods, in: Sociological Methods & Research, 33, No. 3: Special Issue, London: Sage.
- Hummel, H.J., 1972: Probleme der Mehrebenenanalyse. Stuttgart: Teubner.
- Kreft, I.G.G., 1987: Methods and models for the measurement of school effects. Diss., Univ. of Amsterdam. (S)
- Kreft, I.G.G./De Leeuw, J./Kyung-Sung, K., 1990: Comparing four different statistical packages for hierarchical linear regression: GENMOD, HLM, ML2 and VARCL. Center for the Study of Evaluation. CSE Technical Report No. 11.
- Mesarovic, M.D., 1970: Theory of hierarchical, multilevel systems. New York: Academic Press. (S)
- Oosthoek, H./van den Eeden, P. (eds.), 1984: Educational from the multi-level perspective: Models, methodology and empirical findings. New York: Gordon and Breach.
- Pappi, F.U., 1978: Sozialstruktur und politische Konflikte in der BRD: Individual und Kontextanalyse der Wahlentscheidung. Habilitationsschrift, Universität Köln.
- Raudenbush, S.W./Willms, J.D. (eds.), 1991: Schools, classrooms, and pupils: Int. studies of schooling from a multilevel perspective. San Diego: Academic Press. (S)
- Saldern, M. v., 1986 (Hrsg.): Mehrebenenanalyse. Beiträge zur Erfassung hierarchisch strukturierter Realität. Weinheim/München: Beltz.
- Saldern, M. v., 1987: Schulklima von Schulklassen. Mehrebenenanalytische Untersuchungen zu einem ökopyschologischen Konstrukt. Frankfurt a.M.: Peter Lang.

2.9 Kohortenanalyse

- Mason, W.M./Fienberg, S.E., 1985: Cohort analysis in social research. Beyond the identification problem. Berlin u.a.: Springer.

2.10 Stichprobenselektivität

- Wainer, H. (ed.), 1986: Drawing inferences from self-selected samples. New York u.a.: Springer.

2.11 Klassifikationsverfahren/Clusteranalyse

- Anderberg, M.R., 1973: Cluster analysis for applications. New York: Academic Press.
- Bacher, J., 1994: Clusteranalyse. Anwendungsorientierte Einführung. München/Wien: Oldenbourg.
- Bock, H.E., 1974: Automatische Klassifikation. Göttingen: Vandenhoeck & Ruprecht. (S)
- Deichsel, G./Trampisch, H.J., 1985: Clusteranalyse und Diskriminanzanalyse. Stuttgart: G. Fischer.
- Eckes, T./Rosbach, H., 1980: Clusteranalysen. Stuttgart: Kohlhammer.
- Everitt, B.S., 1980: Cluster analysis, 2nd ed. London: Heinemann.
- Everitt, B.S., 1993: Cluster analysis. New York: Halsted Press.
- Gordon, A.D., 1981: Classification: Methods for the exploratory analysis of multivariate data. London: Chapman & Hall.
- Hand, D.J., 1981: Discrimination and Classification. Chichester: Wiley.
- Hartigan, J.A., 1975: Clustering algorithms. New York: Wiley.
- Hudson, H.C., 1982: Classifying social data. San Fransisco: Jossey-Bass.
- Jambu, M., 1983: Cluster analysis and data analysis. Amsterdam: North-Holland.
- Lorr, M., 1983: Cluster analysis for social scientists. San Fransisco: Jossey-Bass.
- Mezzich, J.E., 1980: Taxonomy and behavioral research. Comparative performance of grouping methods. London: Academic Press.
- Rollet, B./Bartram, M., 1976: Einführung in die hierarchische Clusteranalyse: für Psychologen, Pädagogen und Soziologen. Stuttgart: Klett.
- Romesburg, H.C., 1984: Cluster analysis for researchers. Belmont, Calif.: Lifetime Learning Publication.
- Schlosser, O., 1976: Einführung in die sozialwissenschaftliche Zusammenhangs-analyse. Reinbeck: Rowohlt.
- Sneath, P.H.A./Sokal, R.R., 1973: Numerical taxonomy. San Francisco.
- Sodeur, W., 1974: Empirische Verfahren zur Klassifikation. Stuttgart: Teubner.
- Späth, H., 1977: Cluster – Analyse – Algorithmen zur Objektklassifizierung und Datenreduktion, 2. A. München: Oldenbourg.
- Späth, H., 1980: Fallstudien Cluster-Analyse. München: Oldenbourg.
- Steinhausen, D./Langer, K., 1977: Clusteranalyse. Berlin: Walter de Gruyter.
- Tyron, R.C./Bailey, D.E., 1970: Cluster analysis. New York: McGraw-Hill.
- Vogel, F., 1975: Probleme und Verfahren der numerischen Klassifikation. Göttingen. (S)

2.12 Diskriminanzanalyse

- Cacoullos, T. (ed.), 1973: Discriminant analysis and applications. New York: Academic Press.

- Deichsel, G./Trampisch, H.J., 1985: Clusteranalyse und Diskriminanzanalyse. Stuttgart: G. Fischer.
- Eisenbeis, R.A., 1972: Discriminant analysis and classification procedures: Theory and applications. Lexington: Heath.
- Goldstein, M./Dillon, W., 1978: Discrete discriminant analysis. New York: Wiley.
- Hand, D.J., 1981: Discrimination and Classification. Chichester: Wiley.
- Lachenbruch, P.A., 1975: Discriminant analysis. New York: Hafner Press.

2.13 Faktorenanalyse/Kanonische Korrelationsanalyse

- Arminger, G., 1979: Faktorenanalyse. Stuttgart: Teubner. (S)
- Bartholomew, D.J., 1987: Latent variable models and factor analysis. New York: Oxford Univ. Press.
- Chatfield, C./Collins, A.J., 1980: Introduction to multivariate analysis. Chapman & Hall.
- Comrey, A.L., 1992: A first course in factor analysis, 2nd ed. Hillsdale, NY: Erlbaum
- Cureton, E.E./Agostino, R.B.D., 1983: Factor analysis: An applied approach. Hillsdale, NY: Erlbaum.
- Geider, F.J./Rogge, K.-E., 1982: Einstieg in die Faktorenanalyse. Heidelberg: Quelle & Meyer.
- Gorsuch, R.L., 1983: Factor analysis, 2nd ed. Hillsdale, NY: Erlbaum.
- Harman, H.H., 1976: Modern factor analysis. 3rd rev. ed. Chicago: Univ. of Chicago Press.
- Jackson, D.S./Borgatta, E.F., (eds.), 1981: Factor analysis and measurement in sociological research. London: Sage.
- Jolliffe, I.T., 1986: Principal Component Analysis. New York: Springer.
- Lawley, D.N./Maxwell, A.E., 1971: Factor analysis as a statistical method. New York: American Elsevier.
- MacDonald, R., 1985: Factor analysis and related methods. Hillsdale, NY: Erlbaum.
- Revenstorf, D., 1976: Lehrbuch für Faktorenanalyse. Stuttgart: Kohlhammer.
- Revenstorf, D., 1980: Faktorenanalyse. Stuttgart: Kohlhammer.
- Röhr, M., 1993: Statistische Strukturanalysen. Stuttgart: G. Fischer.
- Überla, K., 1974: Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. Berlin u.a.: Springer.
- Weiber, R., 1984: Faktorenanalyse. Eine anwendungsorientierte computergestützte Einführung. St. Gallen.
- Yates, A., 1987: Multivariate data analysis: A perspective on exploratory factor analysis. Albany: Univ. of New York Press.

2.14 Latent – Class – Analyse

- Bartholomew, D.J., 1987: Latent variables models and factor analysis. London: Griffin
- Formann, A.K., 1984: Die Latent-Class-Analyse. Weinheim: Beltz.
- Hagenaars, J.A., 1993: Loglinear models with latent variables. Newbury Park: Sage.
- Heinen, T., 1993: Discrete latent variable models. Tilburg: Tilburg University Press.
- Langeheine, R. (ed.), 1988: Latent trait and latent class models. New York: Plenum Press.
- Lazarsfeld, P.E./Henry, N.W., 1968: Latent structure analysis. Boston: Houghton Mifflin Company.
- Nijkamp, P., 1985: Measuring the unmeasurable. Amsterdam: Nijhoff.
- Vermunt, J.K., 1995: Log-linear event history analysis; a general approach with missing data, latent variables, and unobserved heterogeneity. Tilburg: Tilburg Univ. Press.

2.15 Korrespondenzanalyse

- Benzecri, J.P., 1992: Correspondence analysis handbook. Basel: Dekker.
- Crowder, M.J.; Hand, D.J. 1989: Analysis of repeated measures. London: Chapman and Hall.
- Diggle, P.J.; Liang, K.-Y.; Zeger, S.L., 1994: Analysis of Longitudinal Data. Oxford: Clarendon.
- Fricke, D., 1990: Einführung in die Korrespondenzanalyse. Frankfurt a.M.: R.G. Fischer.
- Geer, J.P. Van de, 1993: Multivariate analysis of categorical data: Theory. Newbury Park: Sage.
- Geer, J.P. Van de, 1993: Multivariate analysis of categorical data: Applications. Newbury Park: Sage.
- Greenacre, M., 1984: Theory and applications of correspondence analysis. London: Academic Press.
- Greenacre, M., 1993: Correspondence analysis in practice. London: Academic Press.
- Greenacre, M./Blasius, J. (eds.), 1994: Correspondence analysis in the social sciences: Recent developments and applications. London: Academic Press.
- Heijden, P.G.M. Van der, 1987: Correspondence analysis of longitudinal categorical data. Leiden: DSWO Press.
- Lebart, L./Morineau, A./Warwick, K.M., 1984: Multivariate descriptive statistical analysis: Correspondence analysis and related techniques for large matrices. New York: Wiley.
- Lindsey, J.k. 1993: Models for repeated measurements. Oxford: Clarendon.

Rijckevorsel, J.L./de Leeuw, J. (eds.), 1988: Component and correspondence analysis. New York: Wiley.

2.16 Multivariate Kontingenztabellenanalyse

- Agresti, A., 1984: Analysis of ordinal categorical data. New York: Wiley.
- Agresti, A., 1990: Categorical data analysis. New York: Wiley.
- Aickin, M., 1983: Linear statistical analysis of discrete data. New York u.a.: Wiley. (S)
- Andersen, E.B., 1980: Discrete statistical models with social science applications. Amsterdam: North-Holland. (S)
- Anderson, E.B., 1991: The statistical analysis of categorical data. 2nd ed. Berlin u.a.: Springer. (S)
- Arminger, G./Küsters, U., 1986: Statistische Verfahren zur Analyse qualitativer Variablen. Bergisch-Gladbach: Bundesanstalt für Straßenwesen.
- Bishop, Y.V.V./Fienberg, S.E./Holland, P.W., 1975: Discrete multivariate analysis: Theory and practice. Cambridge, MA: MIT Press. (S)
- Clogg, C.C./Shihadeh, E.S., 1994: Statistical models for ordinal variables. Thousand Oaks: Sage.
- Everitt, B.S., 1992: The analysis of contingency tables, 2nd ed. London u.a.: Chapman & Hall.
- Eye, A. von, 1990: Introduction to configural frequency analysis: The search for types and antitypes in cross-classifications. Cambridge: Cambridge Univ. Press. (S)
- Fienberg, S.E., 1980: The analysis of cross-classified categorical data. 2nd ed. Cambridge, MA: MIT Press.
- Finney, D.J., 1971: Probit analysis, 3rd ed. Cambridge: Cambridge Univ. Press. (S)
- Fingleton, B., 1984: Models for category counts. Cambridge: University Press.
- Forthofer, R.N./Lehnen, R.G., 1981: Public Program Analysis. A new categorical data approach. Belmont, CA: Lifetime Learning Publications.
- Freeman, D.H., 1987: Applied categorical data analysis. New York: Dekker. (S)
- Geer, J.P., 1993: Multivariate analysis of categorical data. London: Sage. (S)
- Gilber, G.N., 1993: Analyzing tabular data. Loglinear and logistic models for social researchers. London: UCL Press.
- Goodman, L.A., 1984: The analysis of cross-classified data having ordered categories. Cambridge, Mass.: Harvard University Press. (S)
- Goodman, L.A., 1978: Analyzing qualitative/categorical data. Log-linear models and latent-structure analysis. Cambridge, Mass.: Abt Books. (S)
- Haberman, S.J., 1978: Analysis of qualitative data. Vol.I. New York: Academic Press. (S)
- Haberman, S.J., 1979: Analysis of qualitative data. Vol.II. New York: Academic Press. (S)

- Hagenaars, J.A., 1990: Categorical longitudinal data. Log-linear panel, trend, and cohort analysis. London: Sage. (S)
- Harder, T., 1975: Daten und Theorie. München: Fink.
- Hout, M., 1983: Mobility tables. Beverly Hills: Sage. (S)
- Kennedy, J.J., 1992: Analyzing qualitative data. New York u.a.: Praeger.
- Krauth, J./Lienert, G.A.: KFA. Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin. Freiburg: Alber. (S)
- Küchler, M., 1979: Multivariate Analyseverfahren. Stuttgart: Teubner.
- Langeheine, R., 1979: Log-lineare Modelle zur multivariaten Analyse qualitativer Daten. München: Oldenbourg.
- Lautsch, E./Weber, S. von, 1990: Die Konfigurationsfrequenzanalyse (KFA) – Methoden und Anwendungen. Berlin: Volk und Wissen. (S)
- Linder, A./Berchtold, W., 1976: Statistische Auswertung von Prozentzahlen. Basel: Birkhäuser.
- Lienert, G.A., 1988: Angewandte Konfigurationsfrequenzanalyse. Frankfurt a.M.: Athenäum. (S)
- Lindsey, J.K., 1989: The analysis of categorical data using GLIM. New York: Springer. (S)
- Reynolds, H.T., 1977: The analysis of cross-classifications. New York: The Free Press.
- Santner, T.J./Duffy, D.E., 1989: The statistical analysis of discrete data. New York: Springer. (S)
- Schwedler, E., 1984: Möglichkeiten und Grenzen multivariater Analysen auf der Grundlage von Kontingenztabellen – Dargestellt am Beispiel einer Untersuchung zum Erwerbsverhalten verheirateter Frauen. Frankfurt a.M.: P. Lang.
- Tutz, G., 1990: Modelle für kategoriale Daten mit ordinalem Skalenniveau. Göttingen: Vandenhoeck & Ruprecht. (S)
- Upton, G.J.G., 1978: The analysis of cross-tabulated data. New York: Wiley.
- Wermuth, N., 1979: Zusammenhangsanalysen medizinischer Daten. Berlin u.a.: Springer. (S)
- Wickens, T.D., 1989: Multiway contingency tables analysis for the social sciences. Hillsdale, NY: L. Erlbaum Associates.

2.17 Logistische Regressionsanalyse/Modelle diskreter Entscheidungen

- Ben-Akiva, M./Lerman, S., 1985: Discrete choice analysis. Cambridge: MIT-Press. (S)
- Collett, D., 1992: Modelling binary data. London u.a.: Chapman & Hall.
- Cox, D.R., 1970: The analysis of binary data. London: Methuen.
- Cramer, J.S., 1991: The logit model for economists. London: Arnold.
- Daganzo, C., 1979: Multinomial probit. New York: Academic Press. (S)
- Domencich, T.A./McFadden, D., 1975: Urban travel demand: A behavioral analysis. Amsterdam: North-Holland.

- Finney, D., 1971: Probit analysis. Cambridge: Univ. Press. (S)
- Hensher, D.A./Johnson, L.W., 1981: Applied discrete-choice modelling. New York: Halsted Press. (S)
- Hosmer, D.H./Lemeshow, S.L., 1989: Applied logistic regression. New York: Wiley.
- Kleinbaum, D.G., 1994: Logistic regression. A self-learning text. New York: Springer.
- Maddala, G.S., 1983: Limited-dependent and qualitative variables in econometrics. Cambridge: Univ. Press. (S)
- Maier, G./Weiss, P., 1990: Modelle diskreter Entscheidungen. Wien/New York: Springer.
- Manski, C.F./McFadden, D., 1981: Structural analysis of discrete data with econometric applications. Cambridge: MIT Press. (S)
- Morgan, B.J., 1992: Analysis of quantal response data. London u.a.: Chapman & Hall. (S)
- Pudney, S., 1989: Modelling the individual choice. The econometrics of corners, kinks and holes. Oxford: Blackwell. (S)
- Ronning, G., 1991: Mikroökometrie. Berlin u.a.: Springer. (S)
- Train, K., 1986: Qualitative choice analysis. Cambridge: MIT Press. (S)
- Urban, D., 1993: Logit – Analyse. Stuttgart u.a.: G. Fischer.

2.18 Verallgemeinerte lineare Modelle

- Aitkin, M. u.a., 1992: Statistical modelling in GLIM. Oxford: Clarendon Press.
- Andreß, H.-J., 1986: GLIM: Verallgemeinerte lineare Modelle. Wiesbaden: Vieweg.
- Decarli, A., (ed.), 1989: Statistical modelling. Berlin: Springer. (S)
- Dobson, A.J., 1983: Introduction to statistical modelling. London/New York: Chapman & Hall.
- Fahrmeir, L., (Hrsg.), 1992: Advances in GLIM and statistical modelling. Berlin: Springer. (S)
- Francis, B.J./Green, M./Payne, C., 1993: The GLIM Manual. Oxford: Oxford University Press.
- Gilchrist, R. u.a., (ed.), 1985: Generalized linear models. Berlin: Springer. (S)
- Healy, M.J.R., 1993: GLIM: An introduction. Oxford: Clarendon Press.
- McCullagh, P./Nelder, J.A., 1989: Generalized linear models. 2nd ed. London: Chapman & Hall. (S)

2.19 Ereignisdatenanalyse

- Andreß, H.-J., 1992: Einführung in die Verlaufsdatenanalyse. Statistische Grundlagen und Anwendungsbeispiele zur Längsschnittanalyse kategorialer Daten. Historical Social Research Supplementheft No. 5. Köln: Zentrum für Historische Sozialforschung.

- Bergstrom, A.R., 1990: Continuous time econometric modelling. Oxford: Oxford Univ. Press. (S)
- Blossfeld, H.P./Hamerle, A./Mayer, K.U., 1986: Ereignisanalyse: Statistische Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften. Frankfurt a. M./New York: Campus Studium.
- Blossfeld, H.-P./Hamerle, A./Mayer, K.U., 1989: Event history analysis. Hillsdale: Lawrence Erlbaum Associates.
- Blossfeld, H.-P./Rohwer, G., 1995: Techniques of event history modeling. New approaches to causal analysis. Mahwah, New York: Lawrence Erlbaum associates.
- Carter, W.H., 1983: Regression analysis of survival data. New York: Dekker. (S)
- Courgeau, D./Lelievre, E., 1992: Event history analysis in demography. Oxford: Clarendon Press. (S)
- Cox, D.R./Lewis, P.A.W., 1966: The statistical analysis of events. London: Chapman & Hall.
- Cox, D.R./Oakes, D., 1988: Analysis of survival data. London: Chapman & Hall.
- Diekmann, A./Mitter, P., 1984: Methoden zur Analyse von Zeitverläufen. Stuttgart: Teubner.
- Diekmann, A./Mitter, P. (eds.), 1984: Stochastic modeling of social processes. Orlando: Academic Press.
- Diekmann, A./Weick, S. (Hrsg.), 1994: Bevölkerungssoziologische Untersuchungen mit den Methoden der Ereignisanalyse. Berlin: Duncker & Humblot.
- Elandt-Johnson, R.C./Johnson, N.L., 1980: Survival models and data analysis. New York: Wiley. (S)
- Hamerle, A./Tutz, G.: Diskrete Modelle zur Analyse von Verweildauern und Lebenszeiten. Frankfurt a.M.: Campus. (S)
- Heckman, J./Singer, B. (eds.), 1979: Longitudinal analysis of labor market data. Cambridge: Cambridge Univ. Press. (S)
- Kalbfleisch, J.D./Prentice, R.L., 1980: The statistical analysis of failure time data. New York: Wiley. (S)
- Klein, J.P. (ed.), 1992: Survival analysis: state of the art (proceedings of the NATO Advances Research Workshop on Survival Analysis and Related Topics, Columbo, Ohio, USA, 23.28 June 1991) Kluwer. (S)
- Lancaster, T., 1990: The econometric analysis of transition data. Cambridge: Cambridge Univ. Press. (S)
- Lawless, J.F., 1982: Statistical models and methods for lifetime data. New York: Wiley. (S)
- Lee, E.T., 1992: Statistical methods for survival data analysis, 2nd ed. New York u.a.: Wiley. (S)
- Mayer, K.U./Tuma, N.B. (ed.), 1990: Event history analysis in life course research. Madison, Wisc.: Univ. of Wisconsin Press. (S)

- Miller, R.G., 1981: Survival analysis. New York: Wiley.
- Namboodiri, K./Suchindran, C.M., 1987: Life table techniques and their applications. New York: Academic Press.
- Nelson, W., 1982: Applied life data analysis. New York: Wiley. (S)
- Parmar, M./Machin, D., 1995: Survival analysis. Chichester: Wiley.
- Trussell, J., (ed.), 1992: Demographic applications of event history analysis. Oxford: Oxford Univ. Press. (S)
- Tuma, N.B./Hannan, M.T., 1984: Social dynamics: Models and methods. New York: Academic Press. (S)
- Vermunt, J.K., 1995: Log-linear event history analysis; a general approach with missing data, latent variables, and unobserved heterogeneity. Tilburg: Tilburg Univ. Press. (S)
- Voges, W. (Hrsg.), 1987: Methoden der Biographie- und Lebenslaufforschung. Opladen: Westdeutscher Verlag.
- Yamaguchi, K., 1991: Event history analysis. Newbury Park, CA: Sage.

2.20 Kausalmodelle/Pfadanalyse

- Aigner, D.J./Goldberger, A.S., 1977: Latent variables in socioeconomic models. Amsterdam: North Holland.
- Andres, J., 1990: Grundlagen linearer Strukturgleichungsmodelle. Frankfurt a.M.: P. Lang.
- Arminger, G./Müller, F., 1990: Lineare Modelle zur Analyse von Paneldaten. Opladen: Westdeutscher Verlag.
- Bagozzi, R., 1980: Causal models in marketing. New York: Wiley.
- Blalock, H.M., Jr., 1961: Causal inferences in nonexperimental research. Chapel Hill: The University of North Carolina Press.
- Blalock, H.M. u.a. (eds.), 1975: Quantitative sociology. International perspectives on mathematical and statistical modeling. New York: Academic Press.
- Blalock, H.M., Jr. 1982: Conceptualization and measurement in the social sciences. Beverly Hills: Sage.
- Blalock, H.M., Jr. (ed.) 1985: Causal models in the social sciences, 2nd ed., New York: Albine Publ. Co.
- Bollen, K.A., 1989: Structural equations with latent variables. New York: Wiley.
- Boomsma, A., 1983: On the robustness of LISREL against small sample size and non-normality. Amsterdam: Sociometric Research Foundation. (S)
- Byrne, B.M., 1989: A primer of LISREL. Basic applications and programming for confirmatory factor analytic models. New York: Springer.
- Dijkstra, T.K., 1985: Latent variables in linear stochastic models: Reflections on »Maximum Likelihood« and »Partial Least Squares« methods. 2nd ed. Amsterdam: Sociometric Research Foundation. (S)
- Duncan, O.D., 1975: Introduction to structural equation models. New York: Academic Press.

- Everitt, B., 1984: An introduction to latent variable models. Cambridge: Cambridge Univ. Press.
- Fornell, C. (ed.), 1982: A second generation of multivariate analysis. Vol. 1: Methods. Vol. 2: Measurement and Evaluation. New York: Praeger. (S)
- Fuller, W.A., 1987: Measurement error models. New York: Wiley. (S)
- Goldberger, A.S./Duncan, O.D. (ed.), 1973: Structural equation models in the social sciences. New York: Seminar Press.
- Güßefeldt, J., 1988: Kausalmodelle in Geographie, Ökonomie und Soziologie: Eine Einführung mit Übungen und Computerprogrammen. Berlin: Springer.
- Hayduk, L.A., 1988: Structural equation modelling with LISREL: Essentials and advances. Baltimore/London: John Hopkins Univ. Press.
- Heise, D., 1975: Causal analysis. New York: Wiley.
- Hermann, D., 1984: Ausgewählte Probleme bei der Anwendung der Pfadanalyse. Frankfurt a.M.: P. Lang. (S)
- Hildebrandt, L./Rudinger, M./Schmidt, P., 1992: Kausalanalysen in der Umweltforschung. Stuttgart: G. Fischer.
- Hodapp, V., 1984: Analyse linearer Kausalmodelle. Bern: Huber.
- Hummell, H.J./Ziegler, R. (Hrsg.), 1976: Korrelation und Kausalität. Bd. 1 – Bd. 3. Stuttgart: Enke.
- Jackson, D.J./Borgatta, F. (eds.), 1981: Factor analysis and measurement in sociological research. Beverly Hills: Sage.
- James, L.R./Mulaik, S.A./Brett, J.M., 1982: Causal analysis: Assumptions, models and data. Beverly Hills: Sage.
- Jöreskog, K.G., 1963: Statistical estimation in factor analysis: A new technique and its foundations. Stockholm: Almqvist and Wiksell.
- Jöreskog, K.G./Sörbom, D. (eds.), 1979: Advances in factor analysis and structural equation models. Cambridge, Mass.: Abt. Books.
- Jöreskog, K.G./Wold, H. (eds.), 1982: Systems under indirect observation: Causality, structure, prediction, Part I and Part II. Amsterdam: North-Holland.
- Jöreskog, K.D./Sörbom, D., 1993: LISREL 8 with PRELIS 2. User's guide. Uppsala: Department of Statistics.
- Kenny, D.A., 1979: Correlation and causality. New York: Wiley.
- Kiene, B., 1982: Statistische Methoden der Ursachenforschung. Frankfurt a.M.: P. Lang.
- Krader, W., 1991: Neuere Entwicklungen linearer latenter Kovarianzstrukturmodelle mit quantitativen und qualitativen Indikatorvariablen. Frankfurt a.M.: P. Lang. (S)
- Kukuk, M., 1991: Latente Strukturgleichungsmodelle und rangskalierte Daten. Konstanz: Hartung-Gorre. (S)
- Luijben, T., 1989: Statistical guidance for model modification in covariance structure analysis. Amsterdam: Sociometric Research Foundation. (S)
- Loehlin, J.C., 1987: Latent variable models. An introduction to factor, path, and structural analysis. Hillsdale, NY: L. Erlbaum Ass.

- McDonald, R.P., 1985: Factor analysis and related methods. Hillsdale, N.Y.: L. Erlbaum.
- Marsden, P.V., (ed.), 1981: Linear models in social research. Beverly Hills: Sage. (S)
- Möbus, C./Schneider, W.: Strukturmodelle für Längsschnittdaten und Zeitreihen: LISREL, Pfad- und Varianzanalysen. Bern: Huber. (S)
- Opp, Karl Dieter, 1976: Schmidt Peter: Einführung in die Mehrvariablenanalyse. Grundlagen der Formulierung und Prüfung komplexer sozialwissenschaftlicher Aussagen. Reinbeck: Rowohlt.
- Pfeifer, A./Schmidt, P., 1987: LISREL: Die Analyse komplexer Strukturgleichungsmodelle. Stuttgart/New York: G. Fischer.
- Saris, W.E., 1984: Causal modelling in nonexperimental research. An introduction to the LISREL approach. Amsterdam: Sociometric Research Foundation.
- Weede, E., 1977: Hypothesen, Gleichungen und Daten. Kronberg/Ts.: Athenäum Verlag.
- Ziegler, R., 1972: Theorie und Modell. München: Oldenbourg.

2.21 Analyse von Longitudinaldaten

- Arminger, G./Müller, F., 1990: Lineare Modelle zur Analyse von Paneldaten. Opladen: Westdeutscher Verlag.
- Coleman, J.S., 1981: Longitudinal data analysis. New York: Basic Books.
- Crouchley, R. (ed.), 1985: Longitudinal data analysis. Aldershot: Avebury. (S)
- Crowder, M. J./Hand, D. J. (1989): Analysis of repeated measures. London: Chapman and Hall.
- Dale, A./Davies, R.B. (eds.), 1994: Social & political change. A casebook of methods. London: Sage.
- Diggle, P. J./Liang, K.-Y./Zeger, S. L. (1994): Analysis of longitudinal data. Oxford: Clarendon Press.
- Eye, A. von (ed.), 1990: Statistical methods in longitudinal research. Vol. 1: Principles and structuring change. Vol. 2: Time series and categorical longitudinal data. New York: Academic Press.
- Hagenaars, J.A., 1990: Categorical Longitudinal Data. Log-linear panel, trend, and cohort analysis. Newbury Park: Sage.
- Heckman, J.J./Singer, B. (eds.), 1985: Longitudinal analysis of labour market data. Cambridge: Cambridge Univ. Press. (S)
- Hensher, D.A. (ed.), 1987: Longitudinal data methods. Oxford: Pergamon Press. (S)
- Hsiao, C., 1986: Analysis of panel data. Cambridge: Cambridge Univ. Press. (S)
- Kasprzyk, D. u.a. (eds.), 1989: Panel surveys. New York: Wiley.
- Kessler, R./Greenberg, D.F., 1981: Linear panel analysis. New York: Academic Press.

- Lindsey, J. K. (1993): Models for repeated measurements. Oxford: Clarendon Press.
- Maddala, G.S. (ed.), 1993: The econometrics of panel data. Brookfield: Ashgate Publishing Company. (S)
- Magnusson, D. u.a. (ed.), 1991: Problems and methods in longitudinal research: Stability and change. Cambridge u.a.: Univ. Press.
- Mednick, S.A./Harvey, M./Finello, K.M. (eds.), 1984: Handbook of longitudinal research, Vol. 1. New York: Praeger.
- Möbus, C./Schneider, W., 1986: Strukturmodelle für Längsschnittdaten und Zeitreihen: LISREL, Pfad- und Varianzanalysen. Bern: Huber. (S)
- Müller, H.-G., 1988: Nonparametric regression analysis of longitudinal data. New York: Springer. (S)
- Nesselroade, J./Baltes, P. (eds.), 1979: Longitudinal research in the study of behavior and development. New York: Academic Press.
- Plewis, I., 1985: Analysing change. Chichester: Wiley.
- Raj, B./Baltagi, B.H., (eds.), 1992: Panel data analysis. Heidelberg: Physica-Verlag. (S)
- Schulsinger, F. u.a. (eds.), 1981: Longitudinal research. Boston, Mass.: Nijhoff.
- Uncles, M.D. (ed.), 1988: Longitudinal data analysis. Methods and applications. London: Pion.
- Visser, R.A., 1985: Analysis of longitudinal data in behavioral and social research. Leiden: DSWO Press.
- Wiggins, Lee M., 1973: Panel Analysis. Latent Probability Models for Attitude and Behavior Processes. Amsterdam/London/New York: Elsevier. (S)

2.22 *Ökonometrie*

- Amemiya, T., 1985: Advanced econometrics. Cambridge: Harvard Univ. Press.
- Assenmacher, W., 1991: Einführung in die Ökonometrie, 4., erg. A. München: Oldenbourg.
- Common, M.S., 1980: Ökonometrie. Ein Grundriß für Studenten. Stuttgart/New York: G. Fischer.
- Cuthbertson, K. et al., 1992: Applied econometric techniques. Ann Arbor: Univ. of Michigan Press.
- Dofi, J.L./Adibi, E., 1988: Econometric analysis: An applications approach. Englewood Cliffs, NY: Prentice Hall.
- Doran, H.E., 1989: Applied regression analysis in econometrics. New York: Dekker.
- Fomby, T.B./Hill, R./Johnson, S.R., 1984: Advanced econometric methods. New York: Springer.
- Godfrey, L.G., 1991: Misspecification tests in econometrics. Cambridge: Univ. Press. (S)
- Goldberger, A.S., 1964: Econometric theory. New York: Wiley.

- Goldberger, A. S., 1992: A course in econometrics. Cambridge, Mass.: Harvard Univ. Press.
- Green, W., 1993: Econometric analysis, 2nd ed. New York: Macmillan.
- Gujarati, D.N., 1988: Basic econometrics, 2nd ed. New York: McGraw-Hill.
- Harvey, A.C., 1994: Ökonometrische Analyse von Zeitreihen, 2. A. München/Wien: Oldenbourg. (S)
- Heil, J., 1989: Einführung in die Ökonometrie, 3., umgestaltete und verb. A. München: Oldenbourg.
- Hübler, O., 1989: Ökonometrie. Stuttgart: G. Fischer.
- Johnson, A.C./Johnson, M.D., 1987: Econometrics. Basic and applied. New York: Macmillan.
- Judge, G.G., 1988: Introduction to the theory and practice of econometrics, 2nd ed. New York: Wiley.
- Kennedy, P.E., 1992: A guide to econometrics, 3rd ed. Oxford: Blackwell.
- Kmenta, J., 1986: Elements of econometrics, 2nd ed. New York: Macmillan.
- Krämer, W., 1989: Econometrics of structural change. Heidelberg: Physica-Verlag. (S)
- Leserer, M., 1980: Grundlagen der Ökonometrie. Göttingen: Vandenhoeck & Ruprecht.
- Maddala, G.S., 1992: Introduction to econometrics, 2nd ed. New York: Macmillan.
- Morgan, M.S., 1990: The history of econometric ideas. New York: Cambridge Univ. Press. (S)
- Nakhaeizadeh, G./Vollmer, K.H. (Hrsg.), 1990: Neuere Entwicklungen in der angewandten Ökonometrie. Heidelberg: Physica-Verlag. (S)
- Pindyck, R.S./Rubinfeld, D.L., 1981: Econometric models and economic forecasts. New York: McGraw-Hill.
- Rinne, H., 1976: Ökonometrie. Stuttgart: Kohlhammer.
- Ronning, G., 1991: Mikroökonomie. Berlin u.a.: Springer. (S)
- Schaich, E./Brachinger, H.W., 1990: Studienbuch Ökonometrie. Berlin u.a.: Springer.
- Schneeweiß, H., 1990: Ökonometrie. 4., überarb. A. Heidelberg: Physica-Verlag.
- Spanos, A., 1986: Statistical foundations of econometric modelling. Cambridge: Univ. Press.
- Stewart, J., 1991: Econometrics. New York u.a.: Allan.
- Theil, H., 1971: Principles of econometrics. New York: Wiley.
- Wonnacott, R.J./Wonnacott, T.H., 1970: Econometrics. New York: Wiley.

2.23 Zeitreihenanalyse

- Abraham, B./Ledolter, J., 1983: Statistical methods for forecasting. New York: Wiley.

- Achilles, M./Bendisch, J./Hartkopf, B., 1985: Einführung in die Zeitreihenanalyse mit ARIMA-Modellen. Modelldefinitionen und Lösungsverhalten im Zeitbereich. Bonn: GMD-Stud. (S)
- Anderson, O.D., 1976: Time series analysis and forecasting – the Box-Jenkins approach. London: Butterworths.
- Anderson, O.D. (ed.), 1982: Time series analysis: Theory and practice. Amsterdam: North-Holland. (S)
- Aoki, M., 1987: State space modelling of time series. Berlin u.a.: Springer. (S)
- Bastian, J., 1985: Optimale Zeitreihenprognose: empirische Probleme und Lösungen. Frankfurt a.M.: P. Lang. (S)
- Bloomfield, P., 1976: Fourier analysis of time series: An introduction. New York: Wiley. (S)
- Bohley, P., 1992: Statistik, 5., überarb. u. erg. A. München: Oldenbourg.
- Bowerman, B./O'Connell, R.T., 1993: Forecasting and time series, 3rd ed. Belmont: Duxbury Press.
- Box, G.E./Jenkins, G.M., 1976: Time series analysis: Forecasting and control, revised edition. San Francisco u.a.: Holden-Day. (S)
- Brillinger, D.R./Krishnaiah, P.R. (eds.), 1983: Time series in the frequency domain. Amsterdam: North-Holland. (S)
- Brillinger, D.R. u.a. (eds.), 1992: New directions in time series analysis. Part I. Berlin u.a.: Springer. (S)
- Brillinger, D.R. u.a. (eds.), 1993: New directions in time series analysis. Part II. Berlin u.a.: Springer. (S)
- Brockwell, P.J./Davis, R.A., 1991: Time series: Theory and methods, 2nd ed. New York u.a.: Springer. (S)
- Chatfield, C., 1989: The analysis of time series, 4th ed. London: Chapman & Hall.
- Chatfield, C., 1982: Analyse von Zeitreihen: Eine Einführung. München: Hanser.
- Christensen, R., 1991: Linear models for multivariate time series and spatial data. New York: Springer. (S)
- Cleary, J.P., 1981: The beginning forecaster: the forecasting process through data analysis. Belmont: Lifetime Learning Publications.
- Cleary, J.P., 1982: The professional forecaster. Belmont: Lifetime Learning Publications.
- Cryer, J.D., 1986: Time series analysis. Boston: Duxbury Press.
- Dhrymes, P.J., 1981: Distributed lags: Problems of Estimation and formulation. Amsterdam u.a.: North-Holland. (S)
- Dielman, T.E., 1989: Pooled cross-sectional and time series data analysis. New York: Dekker. (S)
- Dufour, J.M. (ed.), 1993: New developments in time series econometrics. Heidelberg: Physica. (S)
- Engle, R./Granger, C.W.J., 1991: Long-run economic relationships. Oxford: Univ. Press. (S)

- Fishman, G.S., 1969: Spectral methods in econometrics. Cambridge, Mass.: Univ. Press. (S)
- Fomby, T.B./Rhodes, G.F.(eds.), 1990: Advances in econometrics: Co-integration, spurious regression and unit roots. London: JAI Press. (S)
- Freeman, H., 1965: Discrete time systems. New York: Wiley. (S)
- Fuchs, W., 1989: Box-Jenkins-Prognosen der kurzfristigen Produktionsentwicklung. Frankfurt a.M.: P. Lang. (S)
- Fuller, W.A., 1976: Introduction to statistical time series. New York: Wiley.
- Glass, G.V./Wilson, V.L./Gottman, Y.M., 1975: Design and analysis of time-series experiments. Boulder, CO: Associated University Press.
- Gottman, J.M., 1981: Time series analysis. A comprehensive introduction for social scientists. Cambridge: Cambridge University Press.
- Gardner, W., 1988: Statistical spectral analysis: a nonprobabilistic theory. Englewood Cliffs: Prentice Hall. (S)
- Gerster, H.J., 1988: Lange Wellen wirtschaftlicher Entwicklung. Frankfurt a.M.: P. Lang. (S)
- Granger, C.W.J./Hatanaka, M., 1964: Spectral analysis of economic time series. Princeton: Univ. Press. (S)
- Granger, C.W.J., 1986: Forecasting economic time series. San Diego: Academic Press.
- Granger, C.W.J., 1990: Modelling economic series. Oxford: Univ. Press.
- Gröhn, E., 1970: Spektralanalytische Untersuchungen zum zyklischen Wachstum der Industrieproduktion in der Bundesrepublik Deutschland 1950–1967. Tübingen: Mohr. (S)
- Haas, P., 1983: Zustands- und Parameterschätzung in ökonometrischen Modellen Hilfe von linearen Filter-Methoden. Frankfurt a.M.: Hain. (S)
- Haase, K./Lütkepohl, H. u.a., 1992: MulTi. A menu-driven GAUSS programm for multiple time series analysis. Kiel: Insitut für Statistik und Ökonometrie, Universität Kiel, Germany. (S)
- Hamilton, J.A., 1994: Time series analysis. Princeton: Univ. Press. (S)
- Hannan, E.J., 1970: Multiple time-series. New York: Wiley
- Harvey, A.C., 1981: The econometric analysis of time series. Oxford: Philip Allan. (S)
- Harvey, A.C., 1990: Forecasting, structural time series models and the Kalman filter. Cambridge u.a.: Cambridge Univ. Press. (S)
- Harvey, A.C., 1993: Time series models. 2nd ed. Oxford: Philip Allan. (S)
- Harvey, A.C., 1994: Ökonometrische Analyse von Zeitreihen, 2. A. München/Wien: Oldenbourg. (S)
- Hoff, J.C., 1983: A practical guide to Box-Jenkins forecasting. London: Life-time Learning Publications.
- Holden, K./Peel, D.A./Thompson, J.L., 1990: Economic forecasting: an introduction. Cambridge u.a.: Cambridge Univ. Press.
- Holz, A., 1983: Schätzmethoden für konstante und variable LAG-Strukturen. Bonn: M+M Wissenschaftsverlag. (S)

- Hüttner, M., 1986: Prognoseverfahren und ihre Anwendung. Berlin/New York: Walter de Gruyter.
- Hylleberg, S. (ed.), 1992: Modelling seasonality. Oxford: Oxford Univ. Press. (S)
- Janacek, G./Switt, L., 1993: Time series. Englewood Cliffs: Prentice Hall.
- Jenkins, G.M., 1979: Practical experiences with modelling und forecasting time series. St. Helier: G. Jenkins & Partners.
- Kendall, M./Ord, J.K., 1990: Time series, 3rd ed. London: Edward Arnold. (S)
- Kirchen, A., 1988: Schätzung zeitveränderlicher Strukturparameter in ökonomischen Prognosemodellen. Frankfurt a.M.: Athenäum. (S)
- Kirchgässner, G., 1981: Einige neuere statistische Verfahren zur Erfassung kausaler Beziehungen zwischen Zeitreihen. Göttingen: Vandenhoeck & Ruprecht.
- König, H./Wolters, J., 1972: Einführung in die Spektralanalyse ökonomischer Zeitreihen. Meisenheim: Athenäum. (S)
- Koopmans, L.H., 1974: The spectral analysis of time series. New York: Academic Press. (S)
- Leiner, B., 1976: Spektralanalysen ökonomischer Zeitreihen. Einführung in Theorie und Praxis moderner Zeitreihenanalyse. Opladen: Gabler. (S)
- Lütkepohl, H., 1991: Introduction to multiple time series analysis. Berlin u.a.: Springer. (S)
- Maier, H., 1989: Methoden zur Schätzung der Ordnung bei autoregressiven Modellen. Stuttgart/Leipzig: Teubner. (S)
- Makridakis, S./Wheelwright, S.C., 1978: Forecasting: Methods and applications. New York: Wiley.
- McCleary, R./Hay, Jr., R.A., 1980: Applied time series analysis for the social sciences. London: Sage.
- Meier, F. (Hrsg.), 1988: Prozeßforschung in den Sozialwissenschaften. Stuttgart: Fischer.
- Michels, P., 1992: Nichtparametrische Analyse und Prognose von Zeitreihen. Heidelberg: Physica. (S)
- Mills, T.C., 1990: Time series techniques for economists. Cambridge: Cambridge University Press. (S)
- Mills, T.C., 1993: The econometric modelling of financial time series. Cambridge: Univ. Press.
- Naeve, P., 1970: Spektralanalytische Methoden zur Analyse von ökonomischen Zeitreihen. Würzburg: Physica. (S)
- Nazem, S.M., 1990: Applied time series analysis for business and economic forecasting. New York/Basel: Dekker.
- O'Donovan, T.M., 1983: Short-term forecasting. Chichester: Wiley.
- Pandit, S.M./Wu, S.M., 1983: Time series and system analysis with applications. New York: Wiley.
- Pankratz, A., 1983: Forecasting with univariate Box-Jenkins models: Concepts and cases. New York: Wiley.

- Pankratz, A., 1991: Forecasting with dynamic regression models. New York: Wiley.
- Parzen, E. (ed.), 1984: Time series analysis of irregularly observed data. New York u.a.: Springer. (S)
- Priestley, M.B., 1987: Spectral analysis and time series, Vol. I, II. New York: Academic Press. (S)
- Puri, M.L. etc. (Eds.), 1987: New developments in time-series analysis. New York: Wiley. (S)
- Reinsel, G.C., 1993: Elements of multivariate time series. New York: Springer. (S)
- Revenstorf, D., 1979: Zeitreihenanalyse für klinische Daten. Weinheim: Beltz.
- Rost, E., 1987: Regressionsmodelle mit stochastischen Koeffizienten im Kontext der Kalman-Filter-Theorie. Frankfurt a.M.: R.G. Fischer. (S)
- Rottleuthner-Lutter, M., 1986: Evaluation mit Hilfe der Box-Jenkins-Methode. Frankfurt a.M.: P. Lang.
- Rüdel, T., 1989: Kointegration und Fehlerkorrekturmodelle. Heidelberg: Physica. (S)
- Schaps, J., 1983: Zur Verwendung des Kalman-Ansatzes für eine Verbesserung der Prognosegüte ökonometrischer Modelle. Braunschweig: Haag + Herchen. (S)
- Schlittgen, R./Streitberg, B., 1994: Zeitreihenanalyse, 5., völlig überarb. u. erw. A. München/Wien: Oldenbourg. (S)
- Schmidt, R., 1984: Konstruktion von Digitalfiltern und ihre Verwendung bei der Analyse ökonomischer Zeitreihen. Bochum: Studienverlag Brockmeyer. (S)
- Schmitz, B., 1987: Zeitreihenanalysen in der Psychologie. Verfahren zur Veränderungsmessung und Prozeßdiagnostik. Wiesbaden: Dt. Studien-Verlag. (S)
- Schmitz, B., 1989: Einführung in die Zeitreihenanalyse. Modelle, Softwarebeschreibung, Anwendungen. Bern/Stuttgart/Toronto: Huber.
- Schneider, W., 1986: Der Kalmanfilter als Instrument zur Diagnose und Schätzung variabler Parameter in ökonometrischen Modellen. Heidelberg/Wien: Physica. (S)
- Schuhr, R., 1991: Lineare vs. nichtlineare Modelle für univariate Zeitreihen. Frankfurt a.M.: P. Lang. (S)
- Schulte, H., 1981: Statistisch methodische Untersuchungen zum Problem langer Wellen. Königstein/Ts.: Athenäum. (S)
- Spree, R., 1978: Wachstumstrends und Konjunkturzyklen in der deutschen Wirtschaft von 1820 – 1913. Göttingen: Vandenhoeck & Ruprecht. (S)
- Stier, W., 1978: Konstruktion und Design von Digitalfiltern zur Analyse und Prognose ökonomischer Zeitreihen. Opladen: Westdt. Verlag. (S)
- Stier, W., 1980: Verfahren zur Analyse saisonaler Schwankungen in ökonomischen Zeitreihen. Berlin u.a.: Springer. (S)

- Subba, R.T., 1993: Developements in time series analysis. In honour of Maurice B. Priestley. New York: Chapman & Hall. (S)
- Termin, J., 1985: Sozioökonomische Indikatoren und Zeitreihenkomponenten – neuere Ansätze bei der ökonometrischen Modellbildung. Bonn: Centaurus. (S)
- Vandaele, W., 1983: Applied time series analysis. Univariate and multivariate methods. Redwood City etc.: Addison-Wesley.
- Vasko, T. (ed.), 1987: The long wave debate. Berlin u.a.: Springer. (S)
- Weber, K., 1990: Wirtschaftsprognostik. München: Vahlen. (S)
- Wei, W.W., 1990: Time series analysis. Univariate and multivariate methods. Redwood City: Addison-Wesley. (S)
- Wheelwright, S.C./Makridakis, S., 1985: Forecasting methods for management. 4th ed. New York: Wiley.

2.24 Nichtparametrische Statistik

- Büning, H./Trenkler, G., 1978: Nichtparametrische statistische Methoden. Berlin/New York: Walter de Gruyter.
- Gibbons, J.D., 1976: Nonparametric methods for quantitative analysis. New York u.a.: Holt, Rinehart and Winston.
- Graff, J., 1992: Nichtparametrische Statistik in den Sozialwissenschaften. Pfaffenweiler: Centaurus-Verlag.
- Maritz, J., 1981: Distributed-free statistical methods. London: Chapman & Hall.
- Puri, M.L., 1985: Nonparametric methods in general linear models. New York: Wiley.
- Renn, H., 1975: Nichtparametrische Statistik. Stuttgart: Teubner.
- Siegel, S., 1976: Nichtparametrische statistische Methoden. Frankfurt a.M.
- Sprent, P., 1989: Applied nonparametric statistical methods. London: Chapman & Hall.

2.25 Grafische Gestaltung von Zahlen und statistische Grafik

- Chambers, J.M. u.a., 1983: Graphical methods for data analysis. Belmont: Wadsworth. (S)
- Cleveland, W.S., 1985: The elements of graphing data. Pacific Grove: Wadsworth. (S)
- Cleveland, W.S./McGill, M.E. (ed.), 1988: Dynamic graphics for statistics. Belmont: Wadsworth Inc. (S)
- Cleveland, W.S., 1993: Visualizing data. Summit, New Jersey: Hobart Press.
- Dickinson, G.C., 1973: Statistical mapping and the presentation of statistics. London: E. Arnold Ltd.
- Everitt, B.S., 1978: Graphical techniques for multivariate data. London: Heinemann.

- Henry, G.T., 1995: Graphing data. Thousand Oaks: Sage.
- Geßler, J.R., 1993: Statistische Graphik. Basel: Birkhäuser.
- Henschke, K./Nagel, M., 1990: Graphische Auswertung von Daten für Mediziner und Naturwissenschaftler. Berlin: VEB Verlag Volk und Gesundheit.
- Huff, D., 1976: How to lie with statistics. England: Penguin Books.
- Imhof, E., 1973: Thematische Kartographie. Berlin: De Gruyter.
- Kafitz, F./Linder, W., 1974: Schaubilder der Statistik. Berlin: Verlag Neue Wirtschafts-Briefe.
- Koberstein, H., 1973: Statistik in Bildern. Stuttgart: Sammlung Poeschel.
- Krämer, W., 1994: So überzeugt man mit Statistik. Frankfurt/New York: Campus.
- Riedwyl, H., 1987: Grafische Gestaltung von Zahlenmaterial, 3. A. Bern/Stuttgart: Paul Haupt.
- Ritter, H., 1991: PC – Graphik – Programme in der Statistik. Vergleichende Gegenüberstellungen mit Anwendungsbeispielen. Stuttgart: G. Fischer.
- Schaller, T., 1990: Business Grafik: Auswahl – Bewertung – Praxis. Düsseldorf: Sybex.
- Schmid, C.F./Schmid, S.E., 1979: Handbook of graphic presentation. New York u.a.: Wiley.
- Schmid, C.E., 1983: Statistical graphics: design, principles, and practices. New York: Wiley.
- Schnell, R., 1994: Graphisch gestützte Datenanalyse. München: Oldenbourg.
- Selby, P.H., 1976: Interpreting graphs and tables. New York: John Wiley.
- Talman, M., 1992: Besser präsentieren mit dem PC. Düsseldorf u.a.: Sybex.
- Toit, S.H.C./Steyn, A.G.W./Stumpf, R.H., 1986: Graphical exploratory data analysis. Berlin u.a.: Springer.
- Tufte, E.R., 1983: The visual display of quantitative information. Cheshire, Conn.: Graphics Press.
- Wang, P.C.C., (ed.), 1978: Graphical representation of multivariate data. New York: Academic Press. (S)
- White, J.W., 1984: Using charts and graphs. New York/London: R.R. Bowker Company.
- Whittaker, J., 1990: Graphical models in applied multivariate statistics. New York: Wiley. (S)
- Zelazny, G., 1989: Wie aus Zahlen Bilder werden: Wirtschaftsdaten überzeugend präsentiert. Wiesbaden: Gabler.

2.26 Sammlungen

Sammlung: »Techniken der empirischen Sozialforschung«

Herausgegeben von Jürgen van Koolwijk und Maria Wicken-Mayser, Oldenbourg Verlag, München/Wien.

Band 1: Methoden der Netzwerkanalyse (1987)

Pappi, F.U.: Einleitung: Die Netzwerkanalyse aus soziologischer Perspektive;
Kappelhoff, P./Ziegler, R.: Teilgruppenbildung in Netzwerken;
Hummell, H.J./Sodeur, W./Kappelhoff, P.: Mikrostrukturen von Gesamtnetzwerken;
Feger, H.: Anwendungsgebiet Sozialpsychologie;
Pappi, F.U./Stelck, K.: Datenanalyse (SONIS).

Band 2: Untersuchungsformen (1975)

Albrecht, G.: Nicht-reaktive Messung und Anwendung historischer Methoden;
Nowotny, H./Knorr, K.K.: Die Feldforschung;
Biervert, B.: Der internationale Vergleich;
Herz, T.A.: Vorhersagestudien;
Aleman, H. von/Ortlieb, P.: Die Einzelfallstudie;
Klingemann, H.D./Mochmann, E.: Sekundäranalyse;
Timaues, E.: Untersuchungen im Laboratorium.

Band 3: Erhebungsmethoden: Beobachtung und Analyse von Kommunikation (1974)

Weidmann, A.: Die Feldbeobachtung;
Manz, W.: Die Beobachtung verbaler Kommunikation im Laboratorium;
Scherer, K.R.: Beobachtungsverfahren zur Mikroanalyse non-verbaler Verhaltensweisen;
Scherer, K.: Ausgewählte Methoden der empirischen Sprachforschung;
Herkner, W.: Inhaltsanalyse;
Mochmann, E.: Automatisierte Textverarbeitung.

Band 4: Erhebungsmethoden: Die Befragung (1974)

Koolwijk, J. van: Die Befragungsmethode;
Kreutz, H./Titscher, S.: Die Konstruktion von Fragebögen;
Erbslöh, E./Wiendieck, G.: Der Interviewer;
Esser, H.: Der Befragte;
Wieken, K.: Die schriftliche Befragung;
Drewe, P.: Methoden zur Identifizierung von Eliten.

Band 5: Testen und Messen (1976)

Besozzi, C./Zehnpfennig, H.: Methodologische Probleme der Index-Bildung;
Huber, H./Schmerkotte, H.: Meßtheoretische Probleme der Sozialforschung;
Lück, H.E.: Testen und Messen von Eigenschaften und Einstellungen;
Wegner, R.: Ratingmethoden;
Betz, D.: Skalierungsverfahren;
Sturm, M./Vajna, T.: Grundzüge der Faktorenanalyse;
Dollase, R.: Soziometrische Verfahren.

Band 6: Statistische Forschungsstrategien (1974)

Helten, E.: Wahrscheinlichkeitsrechnung;
Sturm, M./Vajna, T.: Planung und Durchführung von Zufallsstichproben;
Koolwijk, J. van: Das Quotenverfahren: Paradigma sozialwissenschaftlicher
Auswahlpraxis;
Buttler, G.: Statistische Testverfahren;
Stelzl, I.: Experimentelle Versuchsanordnungen.

Band 7: Datenanalyse (1977)

Höhmman, P./Koolwijk, J. van: Deskriptive Methoden der quantitativen Sozial-
forschung;
Buttler, G.: Demographische Methoden;
Pappi, F.U.: Aggregatdatenanalyse;
Dierkes, M.: Die Analyse von Zeitreihen und Longitudinalstudien;
Allerbeck, K.: Computerunterstützte Datenaufbereitung und Datenanalyse.

Band 8: Kausalanalyse (1986)

Hummell, H.J.: Grundzüge der Regressions- und Korrelationsanalyse;
Jagodzinski, W.: Pfadmodelle mit latenten Variablen: Eine Einführung in das
allgemeine lineare Modell LISREL;
Langeheine, R.: Log-lineare Modelle.

Sammlung: »Die Befragung«

Herausgegeben von Kurt Holm, Francke Verlag, München.

Band 1: Der Fragebogen – Die Stichprobe (1975)

Holm, K.: Zweck und Verlauf einer Befragung;
ders.: Das Modell des Untersuchungsgegenstandes;
ders.: Kausalität;
ders.: Die Frage;
Kirschhofer-Bozenhardt, A. von/Kaplitza, G.: Der Fragebogen;
dies.: Das Interviewernetz;
Kaplitza, G.: Die Stichprobe;
Wilk, L.: Die postalische Befragung.

Band 2: Datenaufbereitung – Tabellenanalyse – Korrelationsmatrix (1975)

Geske, G.: Kodierung der Daten;
Holm, K.: Abholung der Daten und automatische Fehlersuche;
Geske, G.: Eindimensionale Grundauszählung und Normalverteilungstest;
Holm, K.: Anmerkung zu den Computer-Programmen, insbesondere zu den
Algol-Programmen. Liste aller Programme;
ders.: Schritte der Auswertung und Wahl der Auswertungsverfahren;
Schmierer, C.: Tabellenanalyse;
Holm, K.: Einführung in die Matrix-Algebra.

Band 3: Die Faktorenanalyse (1975)

Holm, K.: Die Faktorenanalyse – ihre Anwendung auf Fragebatterien;
ders.: Die Korrelation zwischen zwei Fragebatterien.

Band 4: Skalierungsverfahren – Panelanalyse (1976)

Sixtl, F.: Skalierungsverfahren: Grundzüge und ausgewählte Methoden sozialwissenschaftlichen Messens;

Denz, H.: Likert-Skalierung;

Holm, K.: Die Zuverlässigkeit sozialwissenschaftlichen Messens;

ders.: Die Gültigkeit sozialwissenschaftlichen Messens;

Arminger, G.: Anlage und Auswertung von Paneluntersuchungen.

Band 5: Pfadanalyse – Coleman-Verfahren (1977)

Holm, K.: Die Pfadanalyse;

Haiböck, H.: Nicht-lineare multiple Regression;

Stumpf, H.: Das Coleman-Verfahren;

Holm, K.: Pfadanalyse für dichotome Variable (»Coleman-Pfadanalyse«).

Band 6: Das allgemeine lineare Modell (1979)

Holm, K.: Das allgemeine lineare Modell;

Denz, H.: Das Großgamma-Modell (Die Einbeziehung ordinaler Variabler in das allgemeine lineare Modell);

Arminger, G.: Loglineare Modelle zur Analyse des Zusammenhangs zwischen nominalen variablen;

Arminger, G.: Das allgemeine logistische Modell.

Handbuch der empirischen Sozialforschung, Band I (1967)

Herausgegeben von Rene König.

Wir beziehen uns im folgenden auf die Bände 1 bis 4 der Taschenbuchausgabe (3., umgearbeitete und erweiterte Auflage 1974. Stuttgart: Enke).

Band 1: Geschichte und Grundprobleme der empirischen Sozialforschung

Maus, H.: Zur Vorgeschichte der empirischen Sozialforschung;

Albert, H.: Probleme der Wissenschaftslehre in der Sozialforschung;

Zetterberg, H.: Theorie, Forschung und Praxis in der Soziologie;

Scheuch, K.E.: Entwicklungsrichtungen bei der Analyse sozialwissenschaftlicher Daten.

Band 2: Grundlegende Methoden und Techniken der empirischen Sozialforschung. Erster Teil

König, R.: Die Beobachtung;

Scheuch, E.K.: Das Interview in der Sozialforschung;

Nehnevajsa, J.: Analyse von Panel-Befragungen;

Mangold, W.: Gruppendiskussionen;

Nehnevajsa, J.: Soziometrie.

Band 3a: Grundlegende Methoden und Techniken der empirischen Sozialforschung. Zweiter Teil

Scheuch, K.E.: Auswahlverfahren;
Scheuch, K.E./Zehnpfennig, H.: Skalierungsverfahren;
Hofstätter, P.R.: Faktorenanalyse;
Pages, R.: Das Experiment in der Soziologie.

Band 3b: Grundlegende Methoden und Techniken der empirischen Sozialforschung. Dritter Teil

Neurath, P.: Grundbegriffe und Rechenmethoden der Statistik für Sozialwissenschaftler.

Band 4: Komplexe Forschungsansätze

Mayer, K.: Bevölkerungslehre und Demographie;
Hawley, A.H.: Theorie und Forschung in der Sozialökologie;
Arensberg, C.M.: Soziologie der Gemeinde: die Gemeinde als Objekt und Paradigma;
König, R.: Soziologie der Gemeinde: Neuere Strömungen der Gemeindesozio-
logie;
de Vries Reilingh, H.D.: Soziographie;
Heilfurth, G.: Volkskunde;
Szczepanski, J.: Die biographische Methode;
Silbermann, A.: Systematische Inhaltsanalyse;
Eisermann, G.: Soziologie und Geschichte;
Heintz, P.: Interkultureller Vergleich;
Gesamtregister der Bände 1 bis 4.

Sammlung: »Quantitative Applications in the Social Sciences«

Herausgegeben von Michael S. Lewis-Beck, Newbury Park u.a.: Sage.

1. Basic Math and Statistics:

- (108) Basic Math for Social Sciences, *Timothy M. Hagle*.
- (103) Data analysis. An Introduction, *Michael S. Lewis-Beck*.
- (96) Maximum Likelihood Estimation: Logic and Practice, *Scott R. Eliason*.
- (83) Central Tendency and Variability, *Herbert F. Weisberg*.
- (73) Understanding Significance Testing, *Lawrence B. Mohr*.
- (38) Matrix Algebra: An Introduction, *Krishnan Namboodiri*.
- (35) Introduction to Survey Sampling, *Graham Kalton*.
- (16) Exploratory Data-Analysis, *Frederick Hartwig with Brian E. Dearing*.
- (8) Analysis of Ordinal Data, *David K. Hildebrand, James D. Laing & Howard Rosenthal*.
- (7) Analysis of Nominal Data, (Second Edition) *H. T. Reynolds*.
- (4) Tests of Significance, *Roman E. Henkel*.

2. Regression Analysis:

- (93) Regression with Dummy Variables, *Melissa A. Hardy.*
- (92) Understanding Regression Assumptions, *William D. Berry.*
- (79) Regression Diagnostics: An Introduction, *John Fox.*
- (72) Interaction Effects in Multiple Regression, *James Jaccard, Robert Turrisi & Choi K. Wan.*
- (57) Understanding Regression Analysis, *Larry D. Schroeder, David L. Sjoquist & Paula E. Stephan.*
- (51) Stochastic Parameter Regression Models, *Paul Newbold & Theodore Bos.*
- (50) Multiple Regression in Practice, *William D. Berry & Stanley Feldman.*
- (29) Interpreting and Using Regression, *Christopher H. Achen.*
- (22) Applied Regression: An Introduction, *Michael S. Lewis-Beck.*

3. Experimental Design and Methods:

- (98) Random Factors in ANOVA, *Sally E. Jackson & Dale E. Brashers.*
- (84) ANOVA: Repeated Measures, *Ellen R. Girden.*
- (74) Experimental Design and Analysis, *Steven R. Brown & Lawrence E. Melamed.*
- (58) Randomized Response: A Method for Sensitive Surveys, *James Alan Fox & Paul E. Tracy.*
- (1) Analysis of Variance (Second Edition), *Gudmund R. Iversen & Helmut Norpoth.*
- (54) Multivariate Analysis of Variance, *James H. Bray & Scott E. Maxwell.*
- (30) Test Item Bias, *Steven J. Osterlind.*
- (23) Research Designs, *Paul E. Spector.*
- (12) Analysis of Covariance, *Albert R. Wildt & Olli T. Ahtola.*

4. Measurement and Scaling:

- (89) Multiple Comparison, *Larry E. Toothaker.*
- (82) Summated Rating Scale Construction: An Introduction, *Paul E. Spector.*
- (68) Rasch Models for Measurement, *David Andrich.*
- (67) Analyzing Decision Making: Metric Conjoint Analysis, *Jordan J. Louviere.*
- (61) Multiple Comparisons, *Alan J. Klockars & Gilbert Sax.*
- (36) Achievement Testing: Recent Advances, *Isaac I. Bejar.*
- (26) Multiattribute Evaluation, *Ward Edwards & J. Robert Newman.*
- (25) Magnitude Scaling: Quantitative Measurement of Opinions, *Milton Lodge.*
- (24) Unidimensional Scaling, *John P. McIver & Edward G. Carmines.*
- (17) Reliability and Validity Assessment, *Edward G. Carmines & Richard A. Zeller.*
- (15) Multiple Indicators: An Introduction, *John L. Sullivan & Stanley Feldman.*

- (11) Multidimensional Scaling, *Joseph B. Kruskal & Myron Wish.*
- 5. Factor Analysis, Canonical Correlation and Related Topics:*
- (78) Data Theory and Dimensional Analysis, *William G. Jacoby.*
- (69) Principal Components Analysis, *George H. Duntelman.*
- (47) Canonical Correlation Analysis: Uses and Interpretation, *Bruce Thompson.*
- (34) Covariance Structure Models: An Introduction to LISREL, *J. Scott Long.*
- (33) Confirmatory Factor Analysis: A Preface to LISREL, *J. Scott Long.*
- (14) Factor Analysis: Statistical Methods and Practical Issues, *Jae-On Kim & Charles W. Mueller.*
- (13) Introduction to Factor Analysis: What is It and How To Do It, *Jae-On Kim & Charles W. Mueller.*
- (6) Canonical Analysis and Factor Comparison, *Mark S. Levine.*
- 6. Research Practice:*
- (88) Working With Archival Data: Studying Lives, *Glen H. Elder Jr., Eliza K. Pavalko & Elisabeth C. Clipp.*
- (85) Processing Data: The Survey Example, *Linda B. Bourque & Virginia A. Clark.*
- (80) Computer Assisted Interviewing, *Willem E. Saris.*
- (77) Expert Systems, *Robert A. Benfer, Edward E. Brent Jr. & Louanna Furbee.*
- (63) Survey Questions: Handcrafting the Standardized Questionnaire, *Jean M. Converse & Stanley Presser.*
- (53) Secondary Analysis of Variance, *K. Jill Kiecolt & Laura E. Nathan.*
- (59) Meta-Analysis: Quantitative Methods for Research Synthesis, *Frederic M. Wolf.*
- (49) Basic Content Analysis (Second Edition), *Robert Philip Weber.*
- (42) Using Published Data: Errors and Remedies, *Herbert Jacob.*
- 7. Categorical Data Analysis and Logistic Regression:*
- (106) Applied Logistic Regression Analysis, *Scott Menard.*
- (101) Interpreting Probability Models. Logit, Probit and Other Generalized Linear Models, *Tim F. Liao.*
- (97) Ordinal Log-Linear Models, *Masako Ishii-Kuntz.*
- (94) Loglinear Models with Latent Variables, *Jacques A. Hagenaars.*
- (86) Logit Modeling: Practical Applications, *Alfred DeMaris.*
- (75) Longitudinal Research, *Scott Menard.*
- (64) Latent Class Analysis, *Allan L. McCutcheon.*
- (62) Information Theory: Structural Models for Qualitative Data, *Klaus Krippendorff.*
- (45) Linear Probability, Logit, and Probit Models, *John H. Aldrich & Forrest D. Nelson.*

- (32) Measures of Association, *Albert M. Liebetrau*.
- (31) Mobility Tables, *Michael Hout*.
- (20) Log-Linear Models, *David Knoke & Peter J. Burke*.
- (7) Analysis of Nominal Data (Second Edition), *H. T. Reynolds*.
- 8. Time Series Analysis:**
 - (99) Univariate Tests for Time Series Models, *Jeff B. Cromwell, Walter C. Labys & Michel Terraza*.
 - (100) Multivariate Tests for Time Series Models, *Jeff B. Cromwell, Walter C. Labys & Michel Terraza*.
 - (70) Pooled Time Series Analysis, *Lois W. Sayrs*.
 - (21) Interrupted Time Series Analysis, *David McDowall, Richard McCleary, Errol E. Meidinger & Richard A. Hay Jr.*
 - (9) Time Series Analysis: Regression Techniques (Second Edition), *Charles W. Ostrom Jr.*
- 9. Nonparametric Statistics:**
 - (95) Bootstrapping: A Nonparametric Approach to Statistical Inference, *Christopher Z. Mooney & Robert D. Duval*.
 - (91) Nonparametric Measures of Association, *Jean Dickinson Gibbons*.
 - (90) Nonparametric Statistics: An Introduction, *Jean Dickinson Gibbons*.
- 10. Classification Methods**
 - (102) Typologies and Taxonomies, *Kenneth D. Bailey*.
 - (66) Q Methodology, *Bruce McKeown & Dan Thomas*.
 - (65) Three-Way Scaling and Clustering, *Phipps Arabie, J. Douglas Carroll & Wayne S. DeSarbo*.
 - (44) Cluster Analysis, *Mark S. Aldenderfer & Roger K. Blashfield*.
- 11. Longitudinal Research and Event History Analysis:**
 - (105) Causal Analysis with Panel Data, *Steven E. Finkel*.
 - (76) Longitudinal Research, *Scott Menard*.
 - (46) Event History Analysis: Regression for Longitudinal Event Data, *Paul D. Allison*.
 - (18) Analyzing Panel Data, *Gregory B. Markus*.
- 12. Causal Modeling:**
 - (55) The Logic of Causal Order, *James A. Davis*.
 - (37) Nonrecursiv Causal Models, *William D. Berry*.
 - (3) Causal Modeling, *Herbert B. Asher*.
- 13. Multilevel Data Analysis:**
 - (81) Contextual Analysis, *Gudmund R. Iversen*.
 - (10) Ecological Inference, *Laura Irwin Langbein & Allan J. Lichtman*.

14. Microcomputer:

- (52) Using Microcomputers in Research, *Thomas W. Madron, C. Neal Tate & Robert G. Brookshire.*
- (40) Microcomputer Methods for Social Scientists (Second Edition), *Philip A. Schrodt.*

15. Mathematical Models in the Social Sciences:

- (107) Chaos and Catastrophe Theories, *Courtney Brown.*
- (56) Introduction to Linear Goal Programming, *James P. Ignizio.*
- (48) Models for Innovation Diffusion, *Vijay Mahajan & Robert A. Peterson.*
- (43) Bayesian Statistical Inference, *Gudmund R. Iversen.*
- (41) Game Theory: Concepts and Applications, *Frank C. Zagare.*
- (27) Dynamic Modeling: An Introduction, *Robert Huckfeldt, C.W. Kohfeld & Thomas W. Likens.*
- (2) Operations Research Methods: As Applied to Political Science and the Legal Process, *Stuart S. Nagel with Marian Neef.*
- (101) Interpreting Probability Models, *Liao.*

16. Specific Applications:

- (87) Analytic Mapping and Geographic Databases, *G. David Garson & Robert S. Biggs.*
- (71) Analyzing Complex Survey Data, *Eun Sul Lee, Ronald N. Forthofer & Ronald J. Lorimor.*
- (39) Introduction to Applied Demography: Data Sources and Estimation Techniques, *Norfleet W. Rives Jr. & William J. Serow.*
- (28) Network Analysis, *David Knoke & James H. Kuklinski.*
- (5) Cohort Analysis, *Norval D. Glenn.*

17. Discriminant Analysis:

- (19) Discriminant Analysis, *William R. Klecka.*

Die wichtigsten Titel sind inzwischen als Sammelbände zu verschiedenen Oberthemen zusammengefaßt als »International Handbooks of Quantitative Applications in the Social Sciences« (edited by Michael S. Lewis-Beck):

- Vol. 1: Basic Statistics (1993)
- Vol. 2: Regression Analysis (1993)
- Vol. 3: Experimental Design and Methods (1993)
- Vol. 4: Basic Measurement (1994)
- Vol. 5: Factor Analysis and Related Techniques (1994)
- Vol. 6: Research Practice (1994)

Neukonzeption der Lehrinhalte für das Herbstseminar im ZHSF

Die langjährigen Erfahrungen mit den 14-tägigen Grundkursen I und II und den maximal sechstägigen Aufbaukursen zu unterschiedlichen (fortgeschrittenen) Methoden der Datenanalyse bilden den Hintergrund für strukturelle Veränderungen und einer Neufassung der Lehrinhalte. Ein Zeitraum von drei Jahren erscheint heute für eine umfassende Ausbildung in Methoden der Historischen Sozialforschung als viel zu lang. Die in dem vorliegenden Beitrag vorzustellende Neukonzeption ist das Ergebnis von eigenen Seminar-Erfahrungen, Anregungen von Teilnehmern und intern geführten Diskussionen. Der Beitrag bietet eine erste Orientierungshilfe in Form einer gerafften Darstellung der methodischen Inhalte der neuen Ausbildungseinheiten *Grundkurs (GK)* und *Aufbaukurs (AK)*, um vor diesem Hintergrund die vom ZA-ZHSF erstmalig im Herbstseminar 1995 angebotenen Neukonzeptionen einordnen bzw. abgrenzen zu können.

Leitendes Interesse bei der Realisierung des neuen Ausbildungskonzeptes ist es, Fähigkeiten bei den Seminarteilnehmern zu entwickeln, um

1. das Methodenspektrum empirisch-statistischer Erkenntnisgewinnung in seiner Relevanz und Reichweite zu kennen und beurteilen zu können,
2. Entscheidungskriterien zu kennen, welche statistische Methoden im konkreten Fall einer wissenschaftliche Problemlösung angemessen sind,
3. im Rahmen von Forschungsvorhaben die gestellten methodischen Anforderungen kritisch beurteilen und bewältigen zu können,
4. sich an der Einführungsliteratur zu den im Kurs behandelten Methoden orientieren zu können.

Fähigkeit zur wissenschaftlichen Problembearbeitung heißt, wissenschaftliche Erkenntnisse und Erkenntnismethoden – in einer dem jeweiligen empirischen Forschungsprojekt angemessenen Weise – als Problemlösungsstrategien einzubringen und sie so für die forschungspraktische Arbeit nutzbar zu machen. Zu den wissenschaftlichen Methoden der Erkenntnisgewinnung in der Historischen Sozialforschung gehört wesentlich das breite statistische Methodenrepertoire der empirischen Sozialforschung. Daher soll neben der intensiven Vermittlung der Grundlagen in dem neuen Einführungskurs auch ein breites Spektrum multivariater Datenanalyseverfahren in Form von Überblicken vorgestellt werden. Die Verschachtelung von Theorie, Empirie und statistischen Verfahren wird bei der Vermittlung der Grundlagen in Form einer Sekundäranalyse beibehalten. Durch die exemplarische Übernahme einer Studie und der damit verbundenen Übersichtlichkeit der theoretisch-inhaltlichen Problematik wird die

ausführliche Diskussion der jeweiligen Forschungsschritte möglich. Die Sekundäranalyse wird verbunden mit der Auswahl und Anwendung einfacher bis hin zu komplexen Analyseverfahren. Die Sekundäranalyse ist ferner besonders für die Integration der beiden Kurse geeignet: Beide orientieren sich künftig am gleichen Forschungsbeispiel und damit am gleichen praktischen Forschungsprozeß. Quantitative Methoden werden somit in einem gemeinsamen inhaltlichen Kontext aufeinander bezogen und aufeinander aufbauend vermittelt. Aus den Begrenzungen 'einfacher' bivariater Analyseverfahren kann der Übergang zum Aufbaukurs mit seinem weiterführenden Methodenspektrum abgeleitet werden. Dabei können die unterschiedlichen praktischen Konsequenzen hinsichtlich der Beantwortung der Forschungsfrage (Reichweite der Aussagen) und der Anwendungsformen von statistischen Verfahren vergleichend demonstriert und durch eigene Forschungsschritte nachvollzogen werden. Für die EDV-technische Umsetzung der Datenanalyse wird einheitlich das Statistikprogramm-Paket 'SPSS für Windows (Version 6.0)' herangezogen.

Der Überblick über die weiterführenden Analyseverfahren ist dagegen nicht an Übungsteile gebunden. Damit ist den Teilnehmern aber zumindest der statistische Methodenpool in seinen Zielsetzungen und Anwendungsmöglichkeiten bekannt. Sie sollen in die Lage versetzt werden, im Rahmen von Forschungsarbeiten die Möglichkeiten der Anwendung spezieller Analysemodelle selbständig zu prüfen. Im Bedarfsfall kann sich ein Teilnehmer zielgerichtet zu einem bestimmten Analyseverfahren im ZA-ZHSF beraten lassen oder in einem speziellen ZHSF-Workshop (s.u.) zum Thema ausbilden lassen. Unter den Oberbegriffen »Zeit« und »Dynamik« werden in dem neu konzipierten Aufbaukurs die grundlegenden Fragestellungen der Zeitreihenanalyse und der Ereignisdatenanalyse in Form eines Überblicks behandelt.

Um hinreichend Zeit für diese Erweiterungen zu schaffen, werden die QUANTUM-Workshops im Rahmen des Herbstseminar-Programms nicht mehr stattfinden. Die thematische (und zeitliche) Ausweitung des Herbstseminar-Kursangebots erfordert künftig für den Statistik-Vorlesungsteil zwei Dozenten.

Alternativ zu den QUANTUM-Workshops bietet das ZA-ZHSF künftig neu konzipierte Workshops von ein bis maximal sechs Tagen Dauer über das Jahr verteilt an. Die inhaltliche Ausrichtung orientiert sich dabei an den folgenden Themenschwerpunkten:

- spezielle multivariate Datenanalyseverfahren (z.B. Zeitreihenanalyse, Klassifikationsverfahren, loglineare Tabellenanalyse etc.),
- neue Entwicklungen auf dem Gebiet der Statistik-Software oder Spezialsoftware (z.B. Datenbankprogramm Kleio, EDV-gestützte Inhaltsanalyse etc.),
- thematisch orientierte Methodik (z.B. Mobilitätsforschung, international vergleichende Gesellschaftsforschung etc.).

Tab. 1: Künftige Struktur des Grundkurses zu den Grundlagen der Methodenlehre für Historiker

Einführung	Wesen der Statistik; Grundbegriffe Merkmale, Merkmalsträger; statistische Masse, Stichprobe	1
Merkmale und ihre Messung	Merkmalsausprägungen, Meßniveaus (Skalentypen und zulässige Aussagen)	
Datenaufbereitung	Von der Datenerhebung zur Datenmatrix; Datenmodifikation (Klassenbildung; Zusammenfassung von Merkmalsausprägungen)	2
Univariate Häufigkeitsverteilungen	Tabellarische Darstellung (absolute Häufigkeiten, Summenfunktion, Prozentanteile, Anteilswerte, Verteilungsfunktion); einfache graphische Darstellungen (Säulen-, Balken-, Kreisdiagramm)	
Maßzahlen zur Beschreibung univariater Verteilungen	Lage-Parameter (Modus, Median, mittlere Abweichung, arithmetisches Mittel, Quartile, Box-Plots); Streuungsmaße (Spannweite, Varianz, Standardabweichung)	2
Bivariate Häufigkeitsverteilung (Kontingenztafel)	Generelle Struktur der bivariaten Tabelle (absolute Zellenhäufigkeiten, Prozentuierungsrichtungen)	6
Quantifizierung von Zusammenhängen und Abhängigkeiten in bivariaten Kontingenztafeln	Grundkonzept der statistischen Beziehung (Prozentsatzdifferenzen); χ^2 -Modell (Zusammenhangsmaße: PHI, Cramer's V, Kontingenzkoeffizient); ordinale Merkmale (Kendall's Tau, Gamma, Somer's d); PRE-Maßzahl (Lambda)	
Beziehung zwischen zwei metrischen Merkmalen (Individualdatenanalyse)	Konzept des Streudiagramms, Korrelationsmaß, Spezialfall: Korrelation zwischen zwei 0/1-kodierten Merkmalen; lineare Einfachregression (Funktionsgleichung, Varianzzerlegung, Bestimmtheitsmaß)	3

Beziehung zwischen einem nominalen Merkmal und einem metrischen abhängigen Merkmal	Varianzanalyse (Modell des Mittelwertvergleichs, Variationszerlegung, Bestimmtheitsmaß η^2)	1
Stichprobenverfahren	Zufallsauswahlverfahren; nichtzufalls-gesteuerte Auswahlverfahren	2
Drittvariablenkontrolle in der Tabellenanalyse	Konzept der statistischen Kontrolle (Analyse von Subpopulationen: konditionale Zusammenhangsanalyse); Typologie von idealtypischen Kausalstrukturen (Scheinbeziehung, Intervention, Suppression, Interaktion, Multikausalität)	1
Weiterführende Datenanalyseverfahren	Modelltypen und Entscheidungsbäume; multiple Regression; Pfadanalyse; Faktorenanalyse; Strukturgleichungsmodelle; Clusteranalyse; Modelle für diskrete abhängige Merkmale (lineares Wahrscheinlichkeitsmodell, logistische Regression, loglineare Tabellenanalyse)	8
EDV-Einführung	Einführung in SPSS für Windows	2
Forschungsstrategie	Der Forschungsprozeß in der Historischen Sozialforschung (generalisierter Forschungsablauf und Untersuchungsbeispiel)	3
Methodik	Methodik der Historischen Sozialforschung am Beispiel einer exemplarischen Forschungsfrage	4
EDV-Übungen	Anbindung der Methodeninhalte an eine exemplarische Forschungsfrage (EDV-gestützte Übungen)	15
Plenumsveranstaltung	Erarbeitung von Abschlußberichten; Diskussion der Ergebnisse aus den einzelnen Übungsgruppen (Teilnehmervorträge)	3

Die Zahlen in der dritten Spalte nennen die Anzahl der benötigten Unterrichtseinheiten (eine Einheit umfaßt 1 1/2 Std.).

Tab. 2: Künftige Struktur des Aufbaukurses zu komplexeren Modellen der multivariaten Datenanalyse

Multiple Regressionsanalyse	Funktionsgleichung; statistisches Kontrollkonzept; Variationszerlegung; multiples Bestimmtheitsmaß; semipartielles Bestimmtheitsmaß; Modellvoraussetzungen; Inferenzstatistik	6
Mehrfache Varianzanalyse	Variationszerlegung; Interaktionskonzept; multiple und semipartielle Bestimmtheit; Inferenzstatistik	2
Regression mit Dummy-Variablen	Grundkonzept von 0/1-kodierten Dummy-Variablen; Regression mit Dummy-Variablen (Interpretation der Regressionskoeffizienten); Modell der Kovarianzanalyse (Dummy-Variablen und metrische unabhängige Merkmale); Interpretation des Interaktionskonzeptes im gemischten Fall	3
Lineares Wahrscheinlichkeitsmodell	Regressionsanalyse mit einer diskreten abhängigen Variablen (0,1-kodiert); metrische und/oder diskrete unabhängige Variablen; Vor- und Nachteile des linearen Wahrscheinlichkeitsmodells	1
Logistisches Regressionsmodell	Diskrete abhängige Variable (0/1-kodiert), metrische und/oder diskrete unabhängige Variablen; Interpretation der Regressionskoeffizienten; Maßzahlen der Modellanpassungsgüte (Pseudo- R^2); Ausblick auf multinomiale Logit-Modelle	2
Ausblick: Analyse von zeitlichen Entwicklungen und Analyse von Ereignissen im Zeitverlauf	Fragestellungen der Zeitreihenanalyse; Fragestellungen der Verlaufsdatenanalyse	4

Methodik	Methodik der Historischen Sozialforschung am Beispiel einer exemplarischen Forschungsfrage	6
EDV-Übungen	Anbindung der Methodeninhalte an eine exemplarische Forschungsfrage (EDV-gestützte Übungen)	12
Plenumsveranstaltung	Erarbeitung von Abschlußberichten; Diskussion der Ergebnisse aus den einzelnen Übungsgruppen (Teilnehmervorträge)	3

Anmerkung: Der Schwerpunkt dieses Kursangebots liegt auf dem allgemeinen linearen Modell und seinen Spezialfällen für unterschiedliche Meßniveau-Konstellationen auf Seiten der unabhängigen wie abhängigen Variablen.

Ehemalige Herbstseminar-TeilnehmerInnen können auch selbst Themen vorschlagen; die Vorschläge werden im ZA-ZHSF gebündelt und im Fall von mehreren Interessenten für ein spezielles Gebiet in Form eines Workshops angeboten.

Die Akzente für den künftigen *Crashkurs* sind:

- EDV-Einführung (SPSS für Windows);
- Merkmale, Merkmalsträger, statistische Masse (Grundgesamtheit) und Stichprobe;
- Meßniveaus (Skalentypen);
- tabellarische Darstellung univariater Verteilungen, Maßzahlen der zentralen Tendenz, Streuungsmaße;
- bivariate Kontingenztabellenanalyse;
- Konzept der statistischen Kontrolle in der dreidimensionalen Tabellenanalyse, idealtypische Beziehungsstrukturen;
- Streudiagramm und Korrelationskoeffizient;
- einfache Regressions- und Varianzanalyse.